Journal of Experimental Psychology

CONTENTS

ARTICLES

(*Continued on inside back cover*)

American Psychological Association

Vol. 102, No. 1 January 1974

The *Journal of Experimental Psychology* publishes original experimental investigations which are judged by the Editors to contribute substantially and significantly toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of general psychological theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Manuscripts. Submit manuscripts in *duplicate* to the Editor: David A. Grant, Department of Psychology, University of Wisconsin, Psychology Building, Charter and West Johnson, Madison, Wisconsin 53706. Manuscripts must be accompanied by an abstract of 100 to 150 words *typed on a separate sheet of paper*. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appear in the *APA Publication Manual*, pp. 11–12.

The Journal offers its authors the option of a "blind" review of their submitted manuscripts under the conditions described in the Special Announcement (1972, Vol. 94, No. 3, p. 325). If the author wishes to have his identity hidden from the reviewer, he should request a blind review and take sole responsibility for preparing his manuscript in such a way as to hide his identity.

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the 1967 revision of the *Publication Manual* of the American Psychological Association. When in doubt about practices of the Journal, authors should examine a recent issue.

Clear, concise exposition is demanded; verbose manuscripts may be returned for revision before being reviewed. Piecemeal experiment-by-experiment reporting is discouraged, and integrated series experiments executed simultaneously (e.g., most doctoral dissertations) *must* be reported in a single paper. Regular papers cannot exceed 20 printed pages. Simple experiments, replications, and supplementary experiments may be published as Short Reports of not more than three printed pages. Reports of negative results and failures to replicate published work are also acceptable if they stand on their own merits as contributions to knowledge.

Normally papers are published in order of their receipt, and the current publication lag is about ten months. All of the cost of an author's alterations in galley proof is charged to the author.

Most of the editorial decisions on manuscripts submitted to the Journal are made by the Editor or one of the Associate Editors. In addition, although listed on the Masthead as Consulting Editors, Norman H. Anderson, Helen Bohmer Daly, Murray Glanzer, and Judith Goggin serve as Special Action Editors. As such, they make the editorial decisions on a limited number of manuscripts rather than serve as reviewers for the editors.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1974: $60 domestic, $61 foreign, $6 single issue. APA member rate: $25. Write to Subscription Section, APA.

Reprints: Authors may order reprints of their articles from the printer when they receive galley proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the followiug: Johnson Associates, Inc., 175 Fifth Avenue, New York, New York 10010; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. Undelivered copies resulting from address changes will not be replaced; second-class mail is not forwarded unless subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Published monthly in two volumes, six issues per volume, by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Washington, D.C., and additional mailing offices.

Journal of
Experimental Psychology

VOL. 102, No. 1 JANUARY 1974

EFFECTS OF MASKING TASKS ON DIFFERENTIAL EYELID CONDITIONING: A DISTINCTION BETWEEN KNOWLEDGE OF STIMULUS CONTINGENCIES AND ATTENTIONAL OR COGNITIVE ACTIVITIES INVOLVING THEM[1]

MICHAEL N. NELSON AND LEONARD E. ROSS[2]

University of Wisconisn—Madison

Two experiments investigated the effects of extraneous masking tasks and the viewing of a silent movie on differential eyelid conditioning. In Experiment I, involving 144 college students, masking tasks and movie viewing both significantly decreased differential conditioning performance. More masking-task than no-masking-task *S*s were classified as unaware of stimulus contingencies on the basis of a postexperimental questionnaire, and unaware *S*s showed significantly less differential responding. The masking tasks and movie also reduced the differential conditioning performance of aware *S*s, however, which suggested that factors other than awareness were important in obtaining the effects. Similar masking-task and movie effects were found in Experiment II in which *S*s were informed of the stimulus contingencies prior to the start of the conditioning. Results of the studies were interpreted as indicating that attentional or cognitive activities involving cue contingencies during the experimental session are important factors which should be differentiated from knowledge of the contingencies per se.

When *S*s in a differential classical conditioning situation are required to perform an extraneous or "masking" task during the conditioning session, their differential conditioning performance is much poorer than in the absence of such a task. This effect has been obtained in electrodermal response (EDR) conditioning studies under a variety of masking-task conditions (e.g., Baer & Fuhrer, 1970; Dawson, 1970), and in eyelid conditioning experiments (e.g., Mayer & Ross, 1969; Ross, Wilcox, & Mayer, 1967) which have used a time-estimation masking task.

Studies utilizing masking tasks with the EDR have generally done so in the context of investigating the relationship of awareness to classical conditioning performance, with the masking task being used to decrease *S*'s awareness of stimulus contingencies. Data from studies such as those cited above suggest that differential classical conditioning of the EDR is related to awareness as defined by postexperimental questionnaire reports, and under some circumstances does not occur in the absence of such awareness.

In contrast, differential eyelid conditioning studies have not directly examined the awareness–differential-conditioning rela-

[1] This research was supported by U.S. Public Health Service Research Grant HD-05653 and Training Grant 5-T01-HDOO-117 from the National Institute of Child Health and Human Development.

[2] Requests for reprints should be sent to Leonard E. Ross, Department of Psychology, University of Wisconsin—Madison, Madison, Wisconsin 53706.

tionship when masking tasks are employed, although previous eyelid conditioning data suggest that awareness factors may be involved in the masking task's effect upon differential conditioning. In this regard, Spence (1966) reported that a light-guessing or probability-learning masking task slows the extinction of the conditioned eyelid response by decreasing *S*'s recognition of the acquisition–extinction transition. Similarly, Spence and Platt's (1967) failure to find a partial reinforcement decrement in acquisition under masking-task conditions was interpreted as reflecting the absence of a cognitive set assumed to be operative as a function of nonreinforcement in standard eyelid conditioning situations.

Given these single-cue conditioning effects, it appears that the reduction in differential eyelid conditioning that has been found under masking-task conditions might be a result of the masking task's effects on *S*'s awareness of the relationships among the positive conditioned stimulus (CS+), negative conditioned stimulus (CS−), and unconditioned stimulus (UCS). One purpose of the present study was to investigate this possibility by using a postexperimental questionnaire to assess *S*'s awareness of the stimulus contingencies.

While the data of eyelid conditioning studies clearly show that a time-estimation masking task reduces differential conditioning performance, there has been no attempt to determine if specific characteristics of the masking task are important in obtaining this effect. In particular, it is possible that the common response (the same time estimations) to both CS+ and CS− required by the time-estimation masking-task procedure might in itself be an important factor in reducing differential conditioning performance. To examine this possibility, the present study compared differential conditioning performance under conditions of (*a*) no masking task, (*b*) the previously used time-estimation task, and (*c*) a modified time-estimation task which required *S* to respond differently to the CS+ and CS−. Both the original and modified masking tasks required 2 time-estimation motor responses (button presses) during every intertrial interval (ITI), but the modified masking task required *S*s to spatially order their estimation responses depending upon whether the CS+ or CS− had been presented on the previous conditioning trial.

A final factor included in the study was the presence or absence of a silent motion picture for *S* to view during the conditioning. Since films have been shown in a number of previous differential eyelid conditioning studies that employed auditory stimuli, it was of interest to determine whether or not viewing a film, an act that does not involve any specific task demands, might function as a masking task in reducing S's awareness of stimulus contingencies and differential conditioning performance.

Experiment I consisted of a study and its replication. The replication was added when it appeared that too few *S*s in the original study were unaware of the stimulus contingencies to make meaningful comparisons of aware and unaware *S*s.

Experiment I

Method

Subjects. One hundred and sixty-five students enrolled in introductory psychology courses at the University of Wisconsin served as *S*s to fulfill part of a course requirement. The data from 21 *S*s were rejected, 19 because of movie projector or programming equipment failure and 2 because of *E* error. The remaining 144 *S*s, 72 males and 72 females, were randomly assigned to groups except for the restriction that each group contain an equal number of males and females. The original study and its replication each consisted of 6 groups of 12 *S*s each.

Apparatus. The experiment was conducted in a semisoundproof 3-room laboratory, with programming and recording equipment located in a control room serving 2 testing rooms. Each *S* sat in an adjustable ophthalmologist's chair facing a movie screen. To make their time-estimation responses, *S*s in masking-task groups pressed buttons located on the ends of the arms of their chairs. Monitor lights in the control room that flashed when the buttons were pressed allowed *E* to check for compliance with instructions.

Pure-tone stimuli were generated by Hewlett-Packard audio oscillators and presented through a Grason-Stadler electronic switch set for fast (≈ 10-μsec.) onset and offset ramps. The tones were delivered binaurally to each *S* through Sharpe HA-10A earphones. Movement of *S*'s right eyelid was re-

corded using standard potentiometer–polygraph apparatus.

Design and procedure. The design of each study was a $3 \times 2 \times 2 \times 2$ factorial with 3 masking-task conditions (masking task, modified masking task, and no masking task), movie vs. no movie, and the 2 counterbalanced factors of frequency of CS+ and sex of *S*. The pure-tone CSs of 800 and 2,100 Hz. were presented for 900 msec. in a continuous white-noise background. The loudness of the background white noise was matched by 4 judges to 65-db. (re .0002 dynes/cm²) free-field white noise. The loudness of the 2 tones was matched by the same judges to a 1,000-Hz. tone 85 db. above the mean threshold. The UCS was a .75-psi air puff of 100-msec. duration delivered through a 1-mm. aperture to *S*'s right cornea. An 800-msec. interstimulus interval and 15-, 20-, and 25-sec. *ITIs* ($M = 20$ sec.) were used. Equal numbers of CS+ and CS− trials were presented in a Gellermann sequence.

The original study and its replication were identical except that masking-task *S*s in the original study received a total of 110 trials with a 1-sentence recorded instruction to cease performing their task between Trials 100 and 101, while no-masking-task *S*s were interrupted with a 1-sentence statement concerning either the silent film or attention to the fixation point. Since the data from Trials 101–110 did not show systematic group differences, they were omitted from further consideration, and *S*s in the replication were given only 100 uninterrupted trials.

Instructions were given to *S*s in 2 parts. The first portion was spoken to *S* as the headset was attached and adjusted. Additional recorded instructions were delivered through *S*'s earphones after *E* had left the testing room. For the movie group, preliminary no-masking-task instructions involved telling *S* that the purpose of the experiment was to measure *S*'s reactions to "certain events" which would occur while *S* watched a silent motion picture. Both modified-masking- and masking-task *S*s were told that the experiment was concerned with time estimation during distraction, and that certain "distracting events" would occur while they were viewing a silent film. No-movie groups were treated identically except that all references to a movie were omitted. Instead of viewing a silent film, *S*s in no-movie groups were asked to fixate a 7.6-cm.-diam. flat-back disk fastened to a movie screen and illuminated by a rectangular patch of diffuse white light produced by a movie projector running without film.

The additional recorded instructions presented after *E* left the testing room began with standard neutral instructions which asked *S* to neither aid nor inhibit his natural responses. Masking-task *S*s were also instructed to judge 2- and 5-sec. time intervals from tone offset on each trial, i.e., 2 sec. after a tone had terminated *S* was to push the button on the right arm of his chair, and 5 sec. after the same tone had terminated *S* was to push the button on the left arm of his chair. Modified-masking-task *S*s were instructed to vary the order in which they pressed the buttons to estimate the 2- and 5-sec. intervals. Half of the modified-masking-task *S*s were instructed to press the right-hand button 2 sec. after the high-frequency tone terminated and to press the left-hand button 5 sec. after the high tone had terminated, but to begin with the left-hand button when performing the same two estimations after the low-frequency tone. The other half of the *S*s were instructed to report the estimations in the reverse spatial order, i.e., left then right for the estimations after the high tone, and right then left after the low tone. After questions had been answered, masking- and modified-masking-task *S*s were given the same recorded pretrial instructions given to no-masking-task *S*s, i.e., that they would now be hearing some of the tones that would occur during the experiment. Three pretrials were then given: 1 800-Hz. tone, 1 2,100-Hz. tone, and 1 unpaired air puff. Modified-masking and masking-task *S*s who initially failed to press their time-estimation buttons in the proper order were corrected by *E* between the first few conditioning trials.

Immediately after the last trial each *S* was given a postexperimental questionnaire.[2] All *S*s were asked the same 12 questions (1 question per page) which differed only slightly among the groups in order to be consistent with the appropriate task and movie condition. The first 2 questions were open ended: The first asked *S* what he thought the purpose of the experiment had been, and the second asked *S* what relationships he had noticed between the various stimuli and other aspects of the experiment. Questions 3–8 required *S* to note, on an 11-point scale, his resistance to the experimental procedures, his anxiety, annoyance, alertness, attention to the movie or the fixation point, and his irritation with the air puff. The next 3 questions asked *S* (*a*) if he noticed a change in the strength of the air puff, (*b*) if he prepared himself for the arrival of the air puff, and (*c*) to estimate the total number of air puffs and tones he had received during the experiment. The last question asked *S* to report, on an 11-point scale, the point during the experiment when he had noticed a relationship between the air puff and the tones. Masking- and modified-masking-task *S*s were given 1 extra question, inserted between Questions 7 and 8, which asked them to report how diligently they had performed their time-estimation tasks. The *S*s were not allowed to modify answers to earlier questions.

Results

A conditioned response was defined as the first pen deflection of at least 2 mm.

[2] The questionnaire used was originally developed by G. A. Kimble, and modified by J. W. Moore. The present study used most of the items reported by Moore to correlate with performance, together with some additional questions developed concerning masking-task conditions.

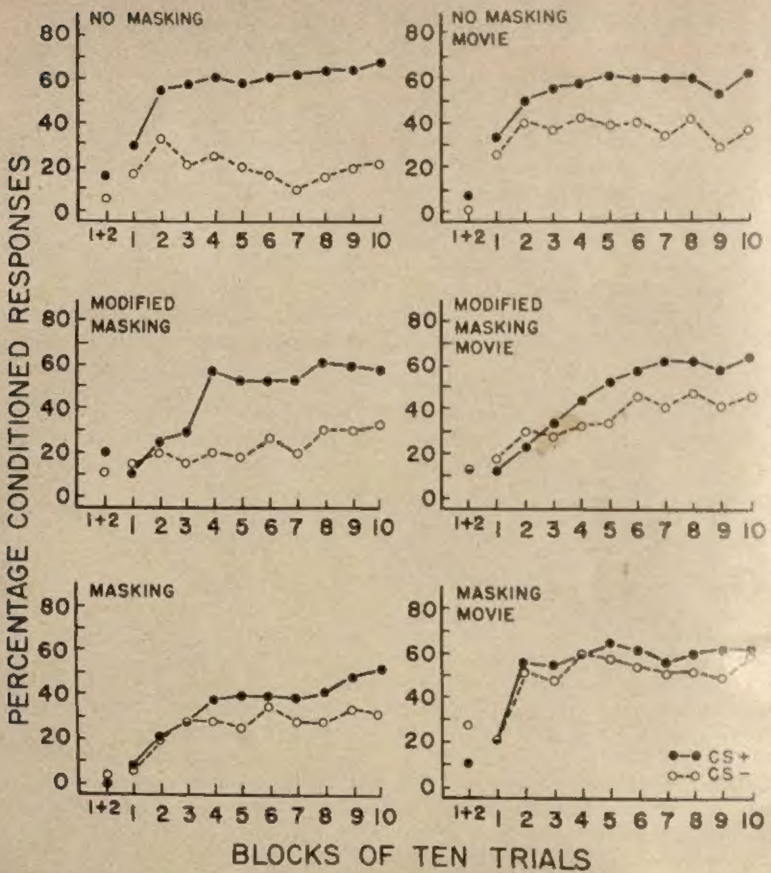


FIGURE 1. Mean percentage responses to the positive conditioned stimulus (CS+) and the negative conditioned stimulus (CS−) as a function of masking-task and movie-viewing conditions collapsed across both replications in Experiment I.

(1 mm. at the eye) in the interval beginning 200 msec. after CS onset and ending at the UCS onset.

Figure 1 presents mean differential responding to CS+ and CS− as a function of task condition and trial block for the movie and no-movie groups of both studies, with the data collapsed across the counterbalanced factors of frequency of CS+ and sex of *S*. Inspection of the figure reveals that differential responding was greatest in the no-masking-task condition, and poorest in the masking-task condition under both movie and no-movie conditions. The movie decreased differential responding under all task conditions.

A 6-factor analysis of variance was performed on difference scores for Trials 1–100, with between-*S*s factors of replications, task condition, movie condition, frequency of CS+, and sex of *S*, and within-*S*s factor of 10-trial blocks. Significant main effects of task condition, F (2, 96) = 10.81, $p < .001$, movie condition, F (1, 96) = 10.23, $p < .005$, and trial blocks, F (9, 864) = 13.61, $p < .001$, were obtained. The replication factor did not reach the .10 level, and the only significant interaction, which proved to be uninterpretable, was that of Replication × Task × Sex × Frequency, F (2, 96) = 4.07, $p < .025$.

Subsequent comparisons of the masking-task conditions revealed that the modified masking-task group was intermediate in performance, i.e., significantly superior t

the masking-task group, F (1, 96) = 4.37, $p < .05$, but inferior to the no-masking-task group, F (1, 96) = 5.39, $p < .025$. The performance of no-masking- and masking-task groups differed significantly, F (1, 96) = 19.48, $p < .001$.

Only 18 of the 144 *S*s were classified as voluntary-form responders by the Hartman and Ross (1961) derivative criterion, which was modified to prevent nonresponders from being classified as voluntary-form responders. Since the data remained virtually unchanged when the voluntary-form-responder data were omitted, separate analyses were not performed.

The postexperimental questionnaire data for the original study and replication were closely examined to determine whether the interruption between Trials 100 and 101 in the original study might have had an effect on *S*'s questionnaire responses. No effect was apparent since *S*s' mean response scores on the questions were quite similar for corresponding groups in the 2 studies. Consequently, *S*s from both studies were classified as aware or unaware of the stimulus contingencies as defined by a correct statement on the first 2 open-ended questions to the effect that the high (or low) tone was followed by a puff of air. Although it proved possible to make the classifications on a quite objective basis, all classification of *S*s as aware or unaware was performed without knowledge of the group assignment or conditioning performance of *S*.

Table 1 presents both the frequencies of *S*s classified as aware and unaware and mean differential responding over all trials for the 6 major group conditions, combined for the study and its replication. It was found that unaware *S*s were distributed unequally across the 6 groups shown in Table 1, χ^2 (2) = 8.02, $p < .02$. The table shows that (*a*) the majority of unaware *S*s were in the 2 original-masking-task groups, (*b*) the modified-masking- and no-masking-task groups did not differ in number of unaware *S*s, and (*c*) the movie had little effect on the distribution of unaware *S*s.

TABLE 1

MEAN PERCENTAGE DISCRIMINATION AS A FUNCTION OF MASKING-TASK AND MOVIE CONDITIONS FOR *S*S CLASSIFIED "AWARE" OR "UNAWARE" ON THE BASIS OF POSTEXPERIMENTAL QUESTIONNAIRE RESPONSES

Movie condition and reported awareness	Task condition		
	Original masking	Modified masking task	No masking task
Movie			
Aware	12.4(9)	11.6(18)	21.6(17)
Unaware	1.3(15)	6.7(6)	12.5(7)
No movie			
Aware	16.5(12)	26.1(17)	40.2(18)
Unaware	.9(12)	19.2(7)	31.3(6)

Note. Number of *S*s in each group is given in parentheses.

A 3 × 2 × 2 × 10 expected cell frequencies analysis of variance involving the factors of task, movie–no movie, awareness, and trial blocks was performed on the CS+/CS− difference score data shown in Table 1. A major finding of the analysis was that *S*s classified as aware of the stimulus contingencies showed significantly greater differential responding than those classified as unaware, F (1, 132) = 6.02, $p < .025$. As in previous analyses, significant main effects of task, F (2, 132) = 7.56, $p < .001$, and movie condition, F (1, 132) = 9.72, $p < .005$, were obtained. The trial blocks effect was significant, F (9, 1188) = 9.89, $p < .001$, but none of the interactions involving the trials variable reached the .10 level. Although task and movie factors did not interact with the awareness variable (both Fs < 1), planned individual comparisons were carried out on masking-task and movie–no-movie differences for aware *S*s only. The no-masking-task–masking-task difference was significant, F (1, 132) = 15.94, $p < .001$, as was the no-masking-task–modified-masking-task difference, F (1, 132) = 8.84, $p < .005$. The masking-task–modified-masking-task difference was not significant, F (1, 132) = 1.12, $p > .20$. The movie–no-movie performance of aware *S*s also differed significantly, F (1, 132) = 13.66, $p < .001$. A similar analysis carried out without the trials factor on terminal differential performance (Trials 61–100) found the same effects significant,

both in the original analysis and individual comparisons.

The data from each group in Table 1 were also examined to determine if significantly more responses were made to CS+ than to CS−. Dependent t tests of the total number of responses to the 2 stimuli revealed greater responding to the CS+ in each aware group ($p < .05$) while no unaware group showed significant differential responding. This was also true for terminal responding (Trials 61–100), except that the difference in responding to CS+ and CS− for aware *S*s tested under movie and original masking conditions only reached the .10 level of significance.

The remaining questions on the postexperimental questionnaire dealing with *S* anxiety, number of tones and air puffs received, annoyance, alertness, attention to task or movie, adaptation to USC, etc., showed nonsignificant low correlations with differential responding; no change in results was noted when the questionnaire data for aware and unaware *S*s were considered separately.

The questionnaire results did not change appreciably when a less specific definition of awareness was used, i.e., when any statement to the effect that the tones were followed by air puffs (CS+ and CS− relationships not necessarily specified) was used as the criterion of awareness.

Discussion

The data of Experiment I clearly demonstrate that performing a time-estimation task or viewing a silent film during the conditioning session results in much poorer differential conditioning performance than would otherwise be the case. Further, the degree to which differential conditioning is reduced can depend upon specific features of the masking task since differential conditioning performance with the original masking task was found to be significantly poorer than that found with the modified-masking task in the analysis involving all *S*s.

The major difference between the masking tasks was the requirement of a common posttrial response as compared to different, cue-contingent, posttrial responses. While the data might be interpreted as indicating that this response difference was responsible for the performance difference, it cannot be determined from the present study whether it was response mediation, a difference in attention to the cues, or some other factor that was the basic mechanism involved. In any case, the data do clearly demonstrate that a masking task can significantly impair differential conditioning performance even when *S* is required to attend to the CS+ and CS− in order to make different responses contingent upon which stimulus is present.

It also should be noted that the masking tasks differed in complexity and undoubtedly imposed different task demands, i.e., cognitive loads, on *S*s. However, comparing the 2 masking-task groups, poorer differential responding was found with the original masking task, which involved considerably less in the way of task demands. Since poorer differential conditioning performance was also found with movie viewing, which required no response on the part of *S*s, the data from this study do not support the position that cognitive load is inversely related to differential conditioning performance. Obviously studies which systematically vary load factors over a wide range of values will be necessary to provide definitive evidence regarding a cognitive-load–differential-conditioning relationship.

A primary purpose of the study was to determine whether the masking-task effect could be accounted for by *S*s' awareness of stimulus contingencies. Since there were more unaware *S*s in the masking-task groups, and unaware *S*s performed significantly more poorly across all conditions, it would appear that the masking task may have prevented some *S*s from becoming aware of the stimulus contingencies and thus affected their differential conditioning performance. Such an effect would be consistent with interpretations of EDR masking-task data (e.g., Dawson, 1970). However, in view of other features of the data, this factor is not sufficient in itself to account for the detrimental effects of the masking task. Most important in this respect is the fact that *S*s who clearly were aware also showed masking-task effects. Similarly, although the movie apparently did not reduce awareness (since the numbers of aware *S*s were very similar under movie and no-movie conditions), aware *S*s who did not view the movie did much better than aware *S*s who did view the movie. Thus, some factor other than *knowledge* of the relationship among the stimuli, as defined by the postexperimental reports of *S*s, would appear

to operate to reduce differential responding under masking-task and movie conditions.

It might be argued that the masking tasks and the movie simply delayed awareness until later in the conditioning session, and that the poorer differential performance of aware masking-task and movie *S*s reflected that delay. This interpretation receives some apparent support from an examination of aware *S*s' responses on the questionnaire item which asked *S* to rate the point at which a relationship between the air puff and the tone had been noticed. A 3 × 2 analysis of variance of the point of awareness data for the masking-task and movie groups did reveal that the no-masking, modified-masking-task, and masking-task group means of 89.4, 84.4, and 67.6 differed significantly, $F\ (2,\ 84) = 6.42$, $p < .005$. The pattern of individual point of awareness differences did not parallel the pattern of differential conditioning performance, however, since subsequent comparisons showed a significant difference between the 2 masking tasks, $F\ (1,\ 84) = 14.26$, $p < .001$, but no significant difference between no masking task and the modified masking task, $F\ (1,\ 84) = 1.82$, $p > .10$: the reverse was true, i.e., significant and nonsignificant differences, respectively, in the case of the conditioning performance comparisons. Further, the point of awareness means for movie and no-movie aware *S*s were quite similar, 81.8 and 83.0, respectively ($F < 1$), although a large and significant difference was found in their differential conditioning performance.

Although these data lead to the conclusion that the point of awareness data do not mirror conditioning differences in a manner that would suggest a causal relationship, it was decided to investigate further the relationship between awareness and differential conditioning in an experiment in which *S*s were informed about the stimulus contingencies *prior* to the start of the conditioning session. Presumably, if the movie and masking-task effects found with aware *S*s were due to reduced or delayed awareness of stimulus contingencies, the movie and masking task should have little effect on performance under conditions where *S* was given this information prior to the conditioning session. On the other hand, decrements similar to those found in Experiment I would support the conclusion that knowledge of stimulus contingencies is not the sole factor responsible for the masking-task and movie-viewing effects on differential conditioning.

Experiment II

Method

Subjects. Fifty-three University of Wisconsin undergraduates enrolled in introductory psychology courses served as *S*s to fulfill part of the course requirement. The data from 5 *S*s were rejected due to movie projector or programming equipment failure. The *S*s were randomly assigned to groups with the restriction that an equal number of males and females were included in each group.

Apparatus. The apparatus remained unchanged from that used in Experiment I, except for the substitution of solid-state timing equipment for Tektronix waveform and pulse generators.

Design and procedure. The design of the experiment was a 2 × 2 × 2 × 2 factorial with 2 task conditions (masking task and no masking task), movie vs. no movie, and counterbalanced factors of sex and frequency of CS+ (800 or 2,100 Hz.). The original masking task was used in this study.

The procedure replicated that of Experiment I except that each *S* was given the following additional instruction prior to the start of conditioning trials: "During the experiment you will be hearing two tones varying in pitch, or frequency. The *high* [low] tone will *always* be followed by a puff of air, while the *low* [high] tone will *never* be followed by a puff of air."

The questionnaire used in Experiment I was also given to *S*s in this experiment.

Results

Mean differential responding data as a function of masking-task and movie conditions are presented in Figure 2 with the data collapsed across the counterbalanced factors of sex and stimulus frequency. Inspection of the figure reveals that the time-estimation masking task and the movie produced fairly large decreases in differential responding relative to the no-movie–no-masking-task group.

A 5-factor analysis of variance was performed on the data of Experiment II, including between-*S*s factors of masking task, movie, sex, and stimulus frequency, and the within-*S*s factor of 10-trial blocks. Performance under the no-masking-task condition was found to be significantly superior to masking-task performance, $F\ (1,\ 32) = 8.06$, $p < .01$, and differential responding increased significantly across trial blocks, $F\ (9,\ 288) = 7.89$, $p < .001$. A significant Movie × Trial Blocks × Frequency interaction was obtained, $F\ (9,\ 288) = 2.15$, $p < .05$. It was not possible to

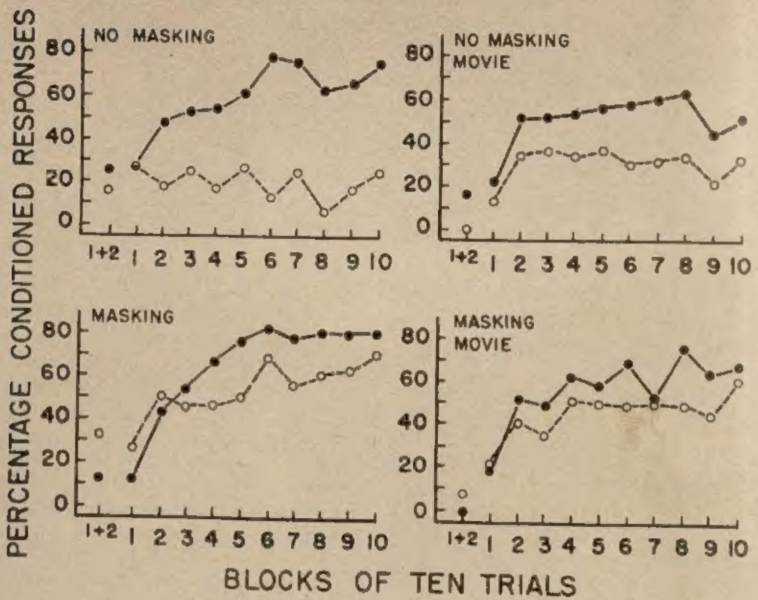


FIGURE 2. Mean percentage responses to the positive conditioned stimulus (CS+) and the negative conditioned stimulus (CS−) as a function of masking-task and movie-viewing conditions in Experiment II.

interpret this interaction in a meaningful way, and a similar interaction was not present in either replication of Experiment I.

Since it had been decided prior to the study to compare the movie–no-movie effect separately for masking and no-masking *Ss*, the analyses were carried out despite the absence of a significant Masking Task × Movie interaction. Under no-masking-task conditions, the movie effect was significant, F (1, 32) = 4.59, $p < .05$, while no effect was found under masking-task-conditions. Inspection of Figure 2 suggests that the masking task so reduced differential responding that movie effects were not possible.

Only 5 *Ss* in Experiment II were classified as voluntary-form responders, and their removal had little effect on the data. Since the questionnaire items were phrased in terms of relationships noticed during the session, *Ss*' answers would be expected to refer to their experiences during the conditioning session rather than to the information given them by *E* prior to the start of the session. The *Ss*' comments during and after the questionnaire suggested that this was the case, but confusion was expressed regarding the apparent contradiction of being informed of the contingencies and then being asked to report when they were noticed. In view of the difficulty in interpreting the questionnaire responses of informed *Ss* of this study, no further analyses of the questionnaire data are reported.

General Discussion

The results of Experiments I and II clearly indicate that the effect of an extraneous task, at least under the experimental conditions of these studies, is not due solely to a concealing or masking of the contingencies from *S* in the sense that *S* is not aware or does not have knowledge of them. It may be that the masking task prevented some *Ss* from becoming aware of the stimulus contingencies so that they could not be reported, and for that matter even without the masking task some *Ss* were not able to report the relationships. It is extremely difficult to determine if *Ss* are unaware in any absolute sense since it is not possible to be certain that *Ss* classified as unaware by one measure would be similarly classified with a different measure. In any case, *Ss* classified as aware in the present

studies performed better than those classified as unaware, and this finding, together with the fact that more masking-task *S*s were classified as unaware, suggests that *part* of the effect of the masking task was to prevent some *S*s from noticing or becoming aware of the cue contingencies so that they could later report them.

In addition, however, aware *S*s from Experiment I, as well as *S*s from Experiment II who were told the contingencies prior to the start of the study, showed reduced differential responding under masking-task and movie conditions. These data strongly suggest that a fundamental factor in differential conditioning is *S*'s attention to the stimulus contingencies, or the result of cognitive activities involving the contingencies (e.g., mediating effects of verbal expectancy), during the conditioning session.

If we assume that some attentional or cognitive activity involving stimulus contingencies enhances differential conditioning, the results of Experiments I and II follow. First, those *S*s who were unaware throughout the study, due to the demands of the masking task or for other reasons, would be less likely to attend to or engage in cognitive activity with respect to cue contingencies and therefore would be expected to perform more poorly. Masking-task or movie *S*s who became aware, or who had knowledge of the contingencies prior to the experiment, would also be less likely to attend to or become cognitively involved with the cue contingencies, but in this case because of the demands of the masking task or the distraction of viewing the movie. Thus, unaware *S*s and aware *S*s under masking-task demands or movie-viewing conditions should both show poorer differential conditioning performance.

Assuming the above to be correct, it appears necessary to distinguish between knowledge of stimulus contingencies, e.g., as measured by postexperimental questionnaires, and *S*s' attentional or cognitive activity involving the relationships among the CS+, CS−, and UCS. It is clear that *S*s may have information about the contingencies, but whether this affects differential conditioning performance undoubtedly depends upon a number of variables such as the particular demands and distracting features of the experimental situation.

Finally, it should be noted that although performance levels varied somewhat across conditions from study to study, the overall picture is one of the masking tasks and movie resulting in a high level of response to the CS−, with good acquisition curves to both CS+ and CS− in almost all groups. Ross and Ross (1972) have discussed the fact that a number of variables affect differential but not single-cue conditioning, and have suggested that single-cue conditioning involves a detection process, while differential conditioning involves detection plus recognition processes. Within this framework the major effects of the masking tasks and movie appear to be on the recognition processes necessary for good differential conditioning rather than upon detection of the CS+ and CS− per se.

REFERENCES

Baer, P. E., & Fuhrer, M. J. Cognitive processes in the differential trace conditioning of electrodermal and vasomotor activity. *Journal of Experimental Psychology*, 1970, **84**, 176-178.

Dawson, M. E. Cognition and conditioning: Effects of masking the CS-UCS contingency on human GSR classical conditioning. *Journal of Experimental Psychology*, 1970, **85**, 389-396.

Hartman, T. F., & Ross, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *Journal of Experimental Psychology*, 1961, **61**, 334-338.

Mayer, M. J., & Ross, L. E. Effects of stimulus complexity, interstimulus interval, and masking task conditions in differential eyelid conditioning. *Journal of Experimental Psychology*, 1969, **81**, 469-474.

Ross, L. E., & Ross, S. M. Conditioned stimulus parameters and the interstimulus inverval: The processing of CS information in differential conditioning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts, 1972.

Ross, L. E., Wilcox, S. M., & Mayer, M. J. A simple eyelid conditioning masking task and its effect on differential conditioning. *Psychonomic Science*, 1967, **9**, 333-334.

Spence, K. W. Cognitive and drive factors in the extinction of the conditioned eye blink in human subjects. *Psychological Review*, 1966, **73**, 445-458.

Spence, K. W., & Platt, J. R. Effects of partial reinforcement on acquisition and extinction of the conditioned eyeblink in a masking situation. *Journal of Experimental Psychology*, 1967, **74**, 259-263.

(Received December 28, 1972)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 10–16

SEEING THE SAME PATTERN TWICE [1]

J. ROSS, J. H. HOGBEN,[2] AND V. Di LOLLO

University of Western Australia, Nedlands, Western Australia, Aust

Two dot patterns, each either brief (3 msec.) or long (250 msec.), were di[illegible] on an oscilloscope one after the other, separated by an interval between 5 and 640 msec. The second pattern was shown in the same field as the first, or in a field which was shifted a small or a large amount to the right. Three *Ss* ju[illegible] whether the second pattern was the same as the first, or different [illegible] radically alters performance for brief patterns by modifying the validit[illegible] criminative visual cues. Performance with long patterns is altered m[illegible] by shift, since *Ss* rely on memory, not on visual effects, under almost all conditions. The results were contrasted to conclusions reached by Pollack.

Pollack (1972) carried out a large study on the detection of changes in spatial position. He made 2 observations about his findings. The first is that the "most striking result from the extended program is related to the sharp maximum in detection performance for interflash intervals (IFIs) of about 64 msec. for brief displays [1972, p. 24]." The second is that "performance is nearly at chance levels for IFIs of 250 msec., which has been taken as an estimate of the duration of the visual image [p. 24]."

At least 3 distinct time ranges are spanned within Pollack's (1972) study: (*a*) a range within which successive patterns appear to be simultaneous (Eriksen & Collins, 1967; Fraisse, 1966; Lichtenstein, 1961; McFarland, 1965); (*b*) a range within which apparent movement is observed between successive patterns or pattern components (Kahneman, 1967); and (*c*) a range beyond apparent movement, within which the first pattern is no longer visible when the second pattern appears. The purpose of the studies reported here is to examine the bases for judgments of pattern identity for dot patterns presented in the same field of view, as in Pollack's study, and for patterns presented in different fields.

If the bases for judgme[illegible] identity are discriminative cues that [illegible] in different time ranges, and if the [illegible]lative location of patterns alters the [illegible]idity of cues available in each range, [illegible]lative location of patterns should govern the effect of temporal parameters such as stimulus duration and interstimulus interval

For example, apparent movement is a valid cue for patterns presented in the same field. Where the second pattern is identical to the first, there will be no apparent movement, since the second set of dots is in registration with the first. However, where the second pattern is different, some dots appear in new positions, and apparent movement will be seen in the appropriate time range. An *O* can conclude validly that the second pattern was different if he sees apparent movement. The same is not true if the second pattern is displaced. Then none of the dots appear in old positions, and apparent movement will be seen both for identical and for different pairs. The movement cue loses its validity if patterns are displaced.

Numerosity is another cue that is valid in one case but not in another. Two identical patterns each of 8 dots seen as simultaneous will be seen as one pattern of 8 dots if the second pattern is not displaced. Two different patterns, on the other hand, will be seen as one pattern of more than 8 dots, since some of the dots in the second pattern occupy new positions. When patterns are displaced, numerosity is no longer a valid cue, since 2 patterns will be seen as one containing

[1] This research was supported by Australian Research Grants Committee (ARGC) Grant ARG. 14.0226.72.MT awarded by the ARGC to J. Ross and V. Di Lollo.

[2] Requests for reprints should be sent to J. H. Hogben, Department of Psychology, University of Western Australia, Nedlands, Western Australia 6009, Australia.

16 dots, whether they are the same or different.

The appearance of 2 different patterns seen as simultaneous is illustrated, for the case of no displacement, in Figure 1. Not only are more dots seen, but there is an obvious "doubling" of some dots, those shifted to new positions to make the second pattern different from the first. This doubling is a prominent and valid cue for patterns that are not displaced. In an effort to destroy its validity, patterns were constructed to look like a pair of different patterns seen as simultaneous (see Figure 1), and were used in Experiment II.

METHOD

Experiment I

Apparatus. The visual display, consisting of 2 fields of random dots, was presented on a gridless Tektronix 602 point plotter equipped with fast P15 phosphor.

The *S* sat inside a lightproof cubicle and viewed the display through a Tektronix (Model 016-0154-00) viewing hood. The scope was positioned just outside the cubicle and the viewing hood was inserted through a light-sealed opening in the wall. The *S* viewed the display binocularly with his face resting against the viewing hood. In that position the display surface was approximately 28 cm. from *S*'s pupils. Low-level illumination of the display surface highlighted a rectangular outline at the periphery of the display area (10 × 8 cm.) and aided *S*'s focusing and convergence.

The dot fields were generated by a PDP-8/S computer, which also performed all the timing and scoring functions. The duration of all timing loops was determined with the aid of a Marconi (Model TF 2414) timer/counter under software control.

Task definition. In all conditions of the experiment the basic display consisted of the presentation of 2 fields each containing a pattern of 8 dots and separated by a variable time interval. On a random half of the trials the configuration of the dots in the second field was the same as that in the first field. On the remaining half of the trials a random 4 of the 8 dots in the second field were displaced to the *S*'s right by 2.5 mm. from the positions they held in the first field. The positioning of the dots within a field was random, subject to the constraint that no 2 dots should appear within 2.5 mm. of each other in either field. The coordinates of all dots in both fields were computed and saved in a table before each trial.

A trial was initiated by the depression of a push button hand-held by the *S*. The *S*'s task was to decide whether the configuration of dots in the 2 fields was or was not the same. He then tapped a telegraph key once or twice to indicate a *same* or *different* response, respectively. The taps were sensed and scored by the computer, which then proceeded to set up the conditions for the next trial. The setting-up procedure took less than 1 sec.

Subjects and procedure. The 3 authors, all experienced *O*s at this type of task, acted as *S*s. It might be noted that although the *O*s were not naive, they used a forced-choice design to overcome the operation of any theoretical and response biases they might have had. Each tried his best to make correct judgments.

Three variables were manipulated in this experiment: duration of the display field, location of the second field with respect to the first on the display surface, and interstimulus interval (ISI). These are considered in turn.

The duration for which a field was displayed was either brief (3 msec.) or long (250 msec.). On brief displays each of the 8 dots was plotted once only. On long displays the 8 dots were plotted sequentially and the total display was continuously refreshed for 250 msec. Brief and long field durations were never mixed within a trial: On any given trial both fields were always of the same duration.

The dots in the first field were chosen randomly on each trial from a 341 × 341 sample space matrix and were plotted on *S*'s left half of the display area within an imaginary square 27 mm. on a side (about 5° 30′ of visual angle). The first field was shown in the same location of the display area under all experimental conditions. The second field was shown in 1 of 3 locations, depending on the experimental condition. In the no-shift condition the second imaginary square was in the same location as the first (see Figure 1), in the small-shift condition it was displaced 5 mm. (see Figure 2), and in the large-shift condition it was displaced 40 mm. to the right of the first (see Figure 3).

The time interval between the 2 dot fields presented on each trial was one of the following: 5, 10, 20, 40, 80, 160, 320, or 640 msec.

The design of the experiment was thus a 2 (Field Durations) × 3 (Field Shifts) × 8 (ISI) factorial with 3 replications in each cell.

Each *S* was tested in 30 experimental sessions of approximately 10-min. duration. In any one session the levels of field duration and field shift were fixed, and 20 consecutive trials were done at each level of ISI. The sequence in which the levels of ISI were presented varied randomly for each session. Within a given session, each of the 8 subsets of 20 trials began with a warning signal (2 dots displayed vertically on the midline of the display area) followed by 4 practice trials not scored for correctness. Twenty experimental trials then followed at the same ISI as the practice trials. Within each block of 20 trials, *same* and *different* pairs occurred with equal probability. The *S*s were

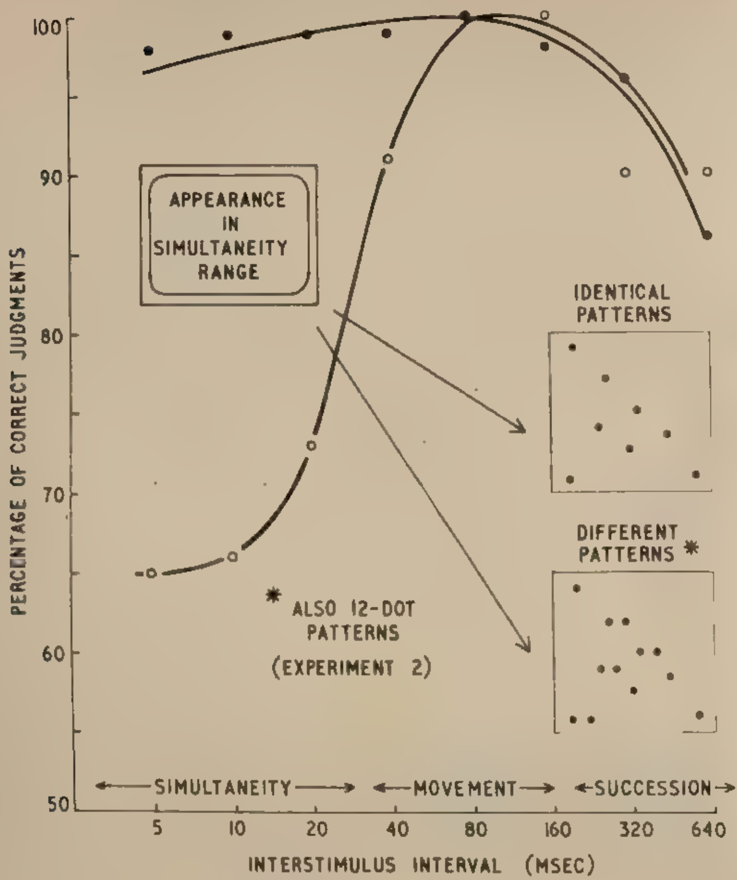


FIGURE 1. Judgment of identity and difference with patterns of brief duration (3 msec.) and no shift of field. (Filled circles show performance when all patterns contain 8 dots, Experiment I. Open circles show performance when some patterns contain 12 dots, Experiment II, and are constructed to resemble the appearance of 2 different 8-dot patterns that are seen as simultaneous. Ranges for simultaneity, apparent movement, and succession without movement are indicated, and the appearance of pairs of patterns within the range of simultaneity is illustrated.)

aware of this fact, which helped to minimize response bias.[3]

Five sessions were run at each of the 6 combinations of field duration and shift. Each *S* worked through the 30 sessions in a sequence of his own choice, making 100 judgments in each of the 48 combinations of the 3 factors.

[3] An analysis of errors for Experiment II shows that errors were predominantly (almost 5 to 1) made by calling different cases *same*, rather than same cases *different*. This may be interpreted to mean either that differences tended to be missed, or that *Ss* adopted the strategy of responding *same* more frequently, when in doubt. No detailed analysis of errors is available for Experiment I.

The *S* received no immediate feedback on the correctness of his responses. However, scores were available at the end of a session, and judgmental strategies and cues were liberally discussed among *Ss* between sessions.

Experiment II

Experiment II differed from Experiment I in 2 respects. First, all displays were presented in the no-shift condition: that is, the second field of dots was always presented within the same imaginary square as the first field. Second, each of the 8 subsets of 20 trials in a given session contained 3 types of displays presented randomly as follows: (*a*) on 25% of the trials the display consisted of

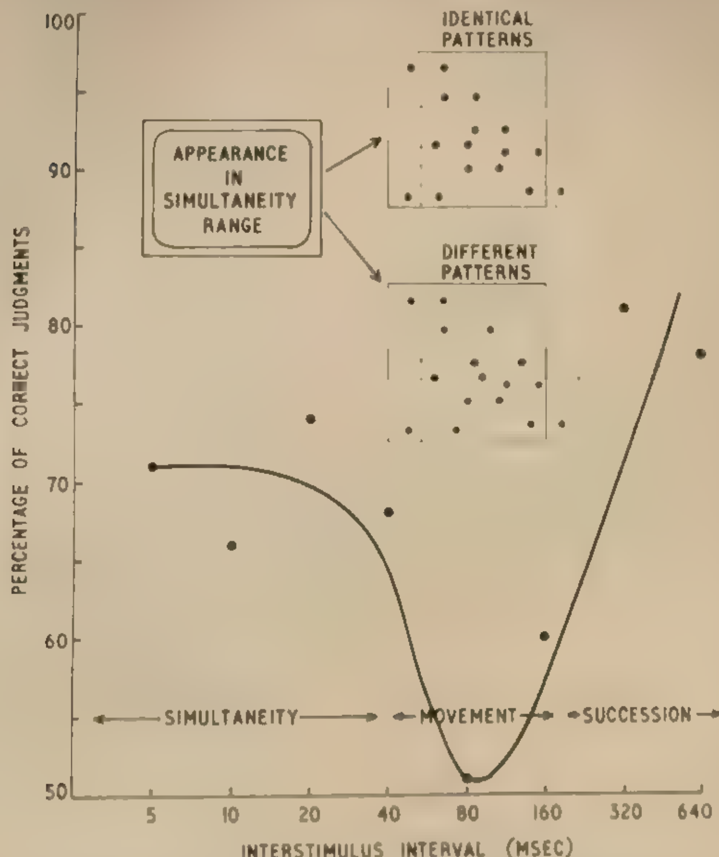


FIGURE 2. Judgment of identity and difference with brief patterns (3 msec.) and small shift (5 mm.) of field. (All patterns contain 8 dots. Experiment I. Ranges for simultaneity, apparent movement, and succession without movement are indicated, and the appearance of pairs of patterns within the range of simultaneity is illustrated.)

8 random dots, which, after the appropriate ISI, reappeared in the same positions on the screen; (*b*) on 50% of the trials a random 4 of the 8 dots were displaced 2.5 mm. to the right of their original positions; (*c*) on the remaining 25% of the trials the display consisted of 8 random dots plus another 4 dots, each of which was plotted 2.5 mm. to the right of one of the other 8 dots. The display therefore consisted of 4 pairs of dots and 4 single dots (see Figure 1). All 12 dots reappeared in the same locations after the appropriate ISI.

As in Experiment I, the *S*'s task was to decide whether the configuration of dots in the 2 fields was or was not the same.

As can be seen, Conditions *a* and *b* replicated the same conditions in Experiment I. Condition *c* was an attempt to simulate the typical appearance assumed by the total display at short ISIs in Condition *b*.

In all other respects, Experiment II was the same as Experiment I. It should be noted that, since Experiment II was run only in the no-shift condition, each *S* was tested in 10 sessions: 5 with displays of brief duration and 5 with displays of long duration.

RESULTS

Brief Patterns

Figure 1 shows the percentage of correct judgments averaged over 3 *O*s, for patterns of brief duration and no shift of field, which is the shift condition used in Pollack's (1972) study. The curve for Experiment I, in which only 8-dot patterns were used, shows high performance

at most ISI values and performance well above chance at an ISI of 640 msec. A sharp maximum in the range 80–160 msec., like that reported by Pollack, can be seen in the curve for Experiment II (Figure 1) in which some patterns contained 12 dots in an arrangement designed to cause confusion with 2 different 8-dot patterns seen as simultaneous.

Figure 2 shows results in the small-shift condition for brief patterns. Performance is at a minimum and almost chance level when ISI is 80 msec., and reaches a maximum at the 2 longest ISI values, 320 and 640 msec. Performance is at an intermediate level for ISI values in the range 5–40 msec.

Figure 3 shows results in the large-shift condition for brief patterns. Performance is at a low level for all ISIs except the longest, 640 msec., and, as in small shift, falls to a minimum, almost chance, level at an ISI of 80 msec.

Long Patterns

Results for all 3 shift conditions with stimuli of long duration are shown in Figure 4. Whereas the amount of shift makes a very marked difference in the shape of the performance curves for brief

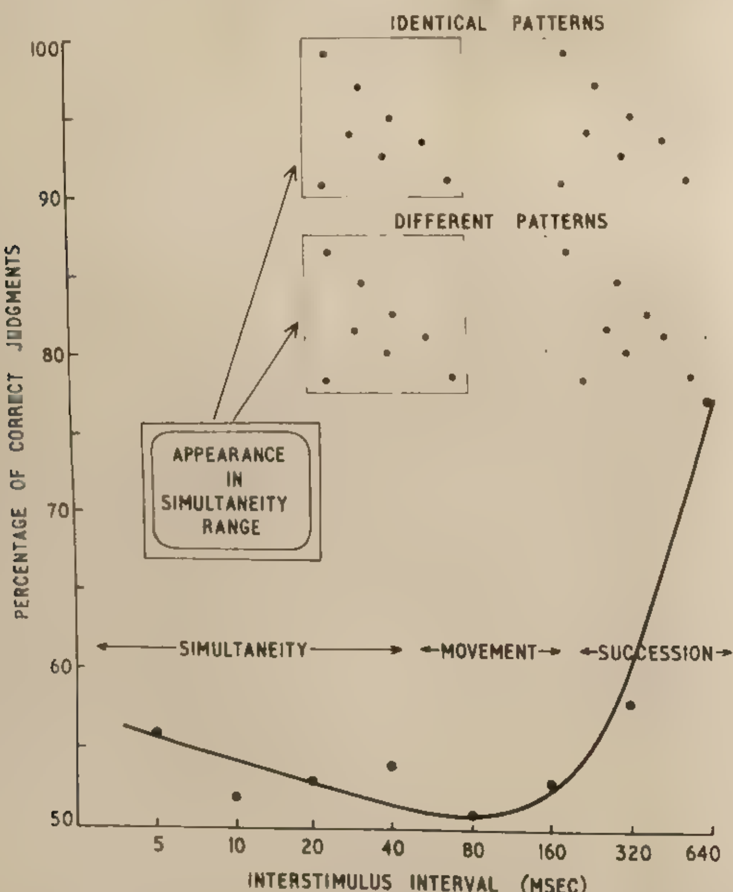


FIGURE. 3. Judgment of identity and difference with brief patterns (3 msec.) and large shift (40 mm.) of field. (All patterns contain 8 dots, Experiment I. Ranges for simultaneity, apparent movement, and succession without movement are indicated, and the appearance of pairs of patterns within the range of simultaneity is illustrated.)

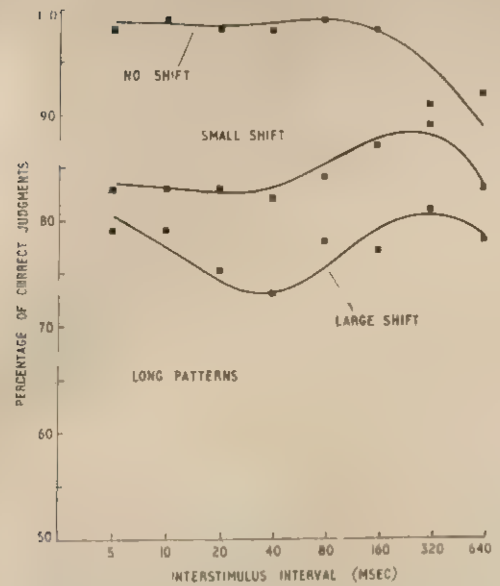


FIGURE. 4. Judgment of identity and difference with patterns of long duration (250 msec.) and no shift, small shift (5 mm.), or large shift (40 mm.) of field.

stimuli, the effect is much less for long stimuli. With increasing shift there is a progressive decline in level of performance, and some change in the shape of performance curves. For both small and large shift, performance is best when ISI is 320 msec., but still high at 640 msec.

A striking feature of the results, illustrated in Figure 5, is that any shift causes performance for brief stimuli separated by an ISI of 80 msec. to drop from near perfect to a chance level. When there is no shift, performance is virtually 100%, but when there is shift, performance is just above 50%. On the other hand, the effect of shift is much less and much the same for brief stimuli separated by the longest ISI, 640 msec., and for long stimuli separated by an ISI of either 80 or 640 msec.

DISCUSSION

Brief Patterns

Consider brief patterns first. In the no-shift case, at ISI values below 40 msec., 2 identical patterns look like 1 pattern of 8 dots, and 2 different patterns like 1 pattern of 12 dots. An *O* can tell the difference on the basis of numerosity or, even more easily, by the doubling illustrated in Figure 1. If, as in Experiment II, the validity of both cues is lowered, the task is harder in this ISI range. There is still a subtle apparent brightness cue, since some dots are plotted only once when patterns are different, but not all *O*s can use it effectively.

When ISI is within the range 40–160 msec., a sequence of different brief patterns with no shift produces apparent movement (Kahneman, 1967) and can readily be discriminated from identical patterns on this basis. At longer ISI values, apparent movement gives way to an apparent succession of patterns both for identical and different patterns, the second now being examined in terms of what is remembered of the first.

In the shift conditions, brief patterns, whether different or not, appear to be simultaneous at ISI values of 40 msec. and below. The small shift allows identical and different patterns to be distinguished on the basis of the equality of distances between pairs of

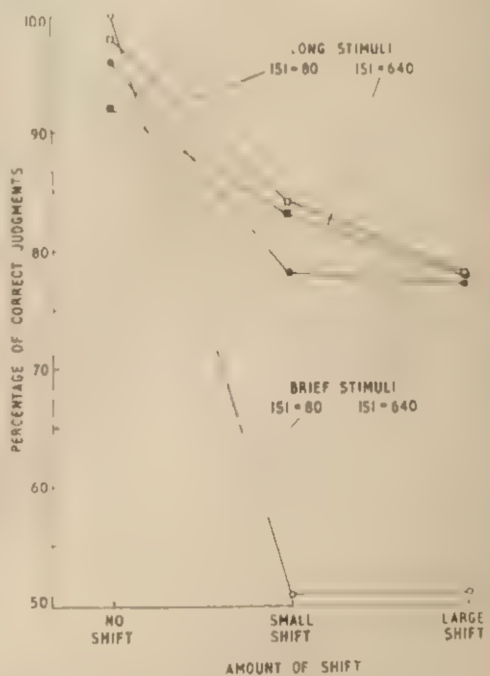


FIGURE 5. Interactions between amount of shift (no shift, 5 mm., or 40 mm.), interstimulus interval (ISI, 80 or 640 msec.), and stimulus duration (3 or 250 msec.).

points (see Figure 2). As the shift is increased to the large value, differential difference becomes more difficult to judge, and performance drops to a low level.

When ISI is between 40 and 160 msec., pronounced apparent movement is observed in both shift conditions for identical as well as different pairs of patterns. What was a discriminative cue in the no-shift condition not only fails to provide a basis for discrimination in both shift conditions, but removes the differential cues available in the lower ISI range.

At long ISI values patterns appear to come in sequence whether there is a shift or not, and performance in all conditions is based on memory, not on discriminative visual effects.

The main features of Pollack's (1972) study are captured in the performance curve for Experiment II (Figure 1) and the top curve of Figure 4. The first shows a sharp maximum in the apparent movement range; the second shows a progressive decline at long ISI values. A comparison of the Pollack-type curve for brief stimuli (Experiment II, Figure 1) with the other curves for brief stimuli (Figures 1, 2, and 3) shows that the Pollack curve is not at all typical. Its shape is explained entirely by the visual effects that are valid as discriminative cues in each of the 3 ranges of ISI values. The different shapes of the other 3 curves can also be explained in exactly the same terms.[4]

Long Patterns

Long patterns are seen as successive at all ISI values, except in the no-shift condition, where, at low ISI values, the points that are not displaced are seen as continuously present, and displaced points as intruding within an established context. Except at the low ISI range in the no-shift condition, there is no obvious visual cue and all comparisons of long patterns can be said to be based on memory. Indeed, the comparison may be based on an encoded version of the first pattern, and the slight rise in performance betwe[illegible]d 160-msec. ISI values may refle[illegible] time required for encoding.

Conclusion

The results of this study do not [illegible]ich contradict Pollack's (1972) observ[illegible] as show that they hold true only unde[illegible]cial circumstances. Performance in com[illegible]ng 2 patterns, or detecting pattern change, [illegible]aximal at ISIs of about 64 msec. and minimal above 250 msec., as Pollack observed, only if patterns are presented in the same field of view, and if the experimental design destroys the validity of numerosity as a cue for brief patterns shown at short intervals. Otherwise, as we have shown, there is no sharp maximum in the apparent movement range. When shift of field is introduced, performance may be minimal at an interval of 80 msec. and maximal at intervals as long as 640 msec.

We believe that Pollack's (1972) results, and ours, fit comfortably with the view that judgments of pattern identity and change rely on visual cues, when they are available and valid. When they are not, the task of comparison is either impossible or becomes a task of memory, comparing features of a pattern seen with what is remembered of a pattern that has been seen but is no longer visible.

[4] It should be noted that at short ISI values for brief stimuli, Pollack's (1972) *S*s would have had available the doubling cue illustrated in Figure 1. We assume that their performance remained low in this range because the experimental design made it difficult for them to discover the cue. Pollack's design is not reported in sufficient detail for us to be certain on this point.

REFERENCES

ERIKSEN, C. W., & COLLINS, J. F. Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology*, 1967, 74, 476–484.

FRAISSE, P. Visual perceptive simultaneity and masking of letters successively presented. *Perception & Psychophysics*, 1966, 1, 285–287.

KAHNEMAN, D. An onset-onset law for one case of apparent motion and metacontrast. *Perception & Psychophysics*, 1967, 2, 577–584.

LICHTENSTEIN, M. Phenomenal simultaneity with irregular timing of components of the visual stimulus. *Perceptual and Motor Skills*, 1961, 12, 47–70.

McFARLAND, J. H. Sequential part presentation: A method of studying visual form perception. *British Journal of Psychology*, 1965, 56, 439–446.

POLLACK, I. Detection of changes in spatial position: Short-term visual memory or motion perception? *Perception & Psychophysics*, 1972, 11, 17–27.

(Received November 16, 1972)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 17–21

EFFECT OF DISTANCE AND SIZE OF STANDARD OBJECT ON THE DEVELOPMENT OF SHAPE CONSTANCY [1]

DALE W. KAESS [2]

University of Aberdeen, Aberdeen, Scotland

S. DZIURAWIEC HAYNES, M. J. CRAIG, S. C. PEARSON, AND J. GREENWELL

Washington State University

The development of shape constancy as a function of the distance, size, and angle of orientation of the standard rectangle was studied. Shape constancy performance improved between 4 and 19 yr., and remained constant to 75 yr. However, a large Age × Distance interaction effect revealed that the age trend occurred with the standard rectangle at 457.20 cm., but not at 91.44 cm., regardless of the size of the object. Apparently shape constancy performance improves from "near" to "far" space as development progresses.

The perception of the physical shape of an object regardless of its orientation from the observer's fronto-parallel plane is commonly referred to as shape constancy performance. Klimpfinger (1933) found that performance on the shape constancy task improves with increasing age from 3 yr. until early adolescence, but then decreases with adults aged 31–70 yr. This apparent inverted-U development trend was replicated by Brault (1962) with data combined from different experiments. Subsequently, however, Meneghini and Leibowitz (1967) have reported a contradictory, linearly decreasing developmental trend with *S*s 4.5–19 yr. of age.

Because Klimpfinger (1933) presented the standard object at a 5-m. observational distance, Meneghini and Leibowitz (1967) supposed that the increasing portion of her obtained age trend might be an artifact of the development of perception at a "far" distance. The "distance hypothesis" suggests that the child first becomes familiar with objects within his reach and only gradually is able to perceive objects at a distance (Piaget, 1967; Vernon, 1962). Wohlwill (1963) has reviewed studies which show that size constancy, or the perception of the size of an object regardless of its distance, is reduced with increased observational distance, and the effect of distance is apparently greater with younger *S*s than with adults. However, there are few data available concerning the parameter of observational distance in shape constancy development.

The only study of the development of shape constancy in which observational distance was systematically varied was reported by Meneghini and Leibowitz (1967). Children 4.5–19 yr. old were shown a rotated 3-cm.-diam. disk at either 91.44- or 457.20-cm. distances, and required to compare it with the comparison series objects presented at a 91.44-cm. distance. Shape constancy performance apparently decreased with age with the test object at a 91.44-cm. distance, but was poor at all ages with the test object at the 457.20-cm. distance. The data were interpreted to demonstrate the validity of the distance hypothesis. However, it is possible that the poor performance by all age groups at the 457.20-cm. distance represents chance performance resulting from the use of too small an object at that distance. The 3-cm. disk shown at 457.20 cm. subtended a visual angle of less than 1°. Thus, the

[1] The authors wish to thank Emma Lou Kenzy, Betty Scott, and Sherril Richardz for making the nursery school subjects available; Patrick Mooney, Principal, and the staff of Edison Elementary School, Pullman, Washington, for providing space for testing 11-year-olds; and staff and members of the Pullman Senior Citizens Center for providing space and data. Also, the assistance of Connie Kaess at every stage of this project is gratefully appreciated.

[2] Requests for reprints should be sent to Dale W. Kaess, Department of Psychology, King's College, University of Aberdeen, Old Aberdeen AB9 2UB, Scotland.

task was probably difficult for the *S*s to perform (Gibson, 1969, p. 365).

Furthermore, since the visual angle subtended by the standard object varied directly with the observational distance, projective size and distance were completely confounded in the Meneghini and Leibowitz (1967) study. That is, the projective size of the standard object was always 5 times larger at the 91.44-cm. distance than it was at the 457.20-cm. distance. Consequently, their results could be interpreted in terms of either observational distance and the distance hypothesis, or in terms of the projective size of the test object as determined by the visual angle subtended by it. This latter interpretation in terms of the size of the test object involves the assumption that acuity for the cues necessary to shape constancy performance varies with the projective size of the object being judged.

Consequently, the present study is designed to investigate the influence of observational distance upon the development of shape constancy performance. Standard objects of 2 sizes have been presented, so that the visual angle subtended by the objects at the 2 distances was systematically varied. Furthermore, elderly *S*s were also used in an attempt to replicate the decreasing trend with older adults, as obtained by Klimpfinger (1933).

Method

Subjects. There were 96 *S*s comprising 4 age groups averaging 4.3, 11.4, 19.5, and 75.6 yr. Each of the 24 *S*s of each age group was randomly assigned to one of 2 treatment combinations denoted by the distance from which the standard object was viewed, so that there were 6 males and 6 females in each group. All *S*s reported having normal binocular vision, at least with corrective lenses.

Apparatus. The standard objects were presented 15 cm. inside a 122 × 91 × 30 cm. flat-black viewing box with a 30-cm.-sq. aperture in the front. The background, 15 cm. behind the object, was also flat black in color. This apparatus was placed on the front edge of a table at *S*'s eye level so that the standard object was either 457.20 or 91.44 cm. from *S*. Comparison objects were presented individually on the background of a similar 122 × 91 × 15 cm. box placed at eye level 91.44 cm. to *S*'s right or left. The *S*s viewed the objects binocularly. The tops of the viewing boxes were open, so that illumination was provided by overhead room lights. There were no visible shadows present on either stimulus.

Stimuli. Two rectangular standard objects were used. They were cut from white textured poster board and mounted upon 2-mm. wires along their longer, vertical dimensions. The standard objects were 7.62 × 8.89 cm. and 38.10 × 44.45 cm., thus preserving the same horizontal-to-vertical ratio. The sizes were chosen so that the larger object would present the same projective size at 457.20 cm. that the smaller object presented at the 91.44-cm. distance. Each of these standard objects was rotated about its vertical axis to either 25° or 75° from the fronto-parallel plane for presentation.

The comparison objects were 2 separate sets of 17 rectangles each with the same height as the corresponding standard object (8.89 or 44.45 cm.), and varying in steps of .63 cm. from 5.04 cm. narrower to 5.04 cm. wider than the corresponding standard object. These comparison objects were also cut from white textured poster board.

Procedure. On an individual trial, one of the 2 standard objects was presented at either a 25° or a 75° orientation with either a 91.44- or a 457.20-cm. observational distance. The rectangles of the comparison series were shown individually in random order 91.44 cm. to *S*'s right or left. The large or small comparison series forms were always presented with the rectangle of the same height. A different random order was used with each trial for each *S*, and order of presentation of trials was counterbalanced across *S*s in each Age × Distance group.

Instructions. All *S*s were given instructions similar to those used by Kaess (1970). It was explained that *S* was participating in a test of how well he could tell the "real" shape of an object even when it is turned to the side. The *S* was shown (by measuring the physical shape of a rotated rectangle) that the real shape remains the same regardless of the orientation. However, no mention was made of the apparently narrower projective shape, or that *S* should compensate for the effects of angle. The youngest *S*s were also told that they were "playing a game," and that they should attempt to be correct every time. No feedback was given during testing, however.

Results

The basic data were the median widths of those comparison series rectangles judged to be equal to each standard rectangle in each series of trials by each *S*. These values were converted to constant error scores by subtracting the width of the standard rectangle from that of the median comparison rectangle judged equal to it, so that the data for the 7.62- and 38.10-cm. rectangles would be comparable (Table 1). The constant errors were subjected to a 4 (age) × 2 (sex) × 2 (distance) × 2 (size)

TABLE 1

CONSTANT ERRORS FROM POINTS OF OBJECTIVE EQUALITY (POINT OF SUBJECTIVE EQUALITY MINUS STANDARD)

		7.62-cm. form		38.10-cm. form	
		25° angle	75° angle	25° angle	75° angle
91.44-cm. distance					
4.2	[illegible]	.05	−.25	.06	−.11
11.5	.38	−.04	−.12	.04	−.03
[illegible]	[illegible]	−.15	−.27	.15	−.02
[illegible]	[illegible]	.05	−.07	−.13	−.17
457.20-cm. distance					
4.4	.31	−.21	−.73	−.27	−.78
11.3	.29	−.29	−.54	−.05	−.30
19.2	2.11	.02	−.28	.28	−.12
75.9	9.57	.01	.03	−.09	−.47

× 2 (angle) analysis of variance, with age, sex, and distance as between-*S*s effects.

The main effect of age, F (3, 80) = 3.48, p < .025, was significantly large. Comparisons of individual means (Neuman-Keuls technique) indicate that the 4-year-olds' mean constant error was significantly larger than that of the 11-year-olds, which was in turn larger than the means of both the 19- and 75-year-olds. Apparently, shape constancy performance improves between 4 and 19 yr. of age, and remains relatively constant with older *S*s.

The effects of distance, F (1, 80) = 13.46, p < .001, and of angle, F (1, 80) = 4.33, p < .05, were also significantly large. The *S*'s judgments of the widths of the rotated forms were more accurate when the forms were viewed from a near, 91.44-cm., distance or when the forms were viewed at 25°, rather than 75°, rotation from their fronto-parallel plane. The Sex × Distance interaction, F (1, 80) = 4.37, p < .05, was also significantly large, however. Comparison of individual means reveals that while the effect of distance obtained with *S*s of both sex, the effect was more pronounced with the male *S*s.

As can be seen in Figure 1, the Age × Distance interaction, F (3, 80) = 4.21, p < .01, was large. Comparisons of individual means reveal that there were no significant differences between age groups with the standard object at the 91.44-cm. distance. But with the standard object at the 457.20-cm. distance, the 4-year-olds' mean constant error was significantly larger than that of the 11-year-olds, which was in turn larger than all others. The mean constant errors of the 19- or 75-year-olds with the 2 observational distances did not differ significantly. Thus, the influence of the distance of the standard object is reduced as age increases, and does not produce a significant difference in the shape constancy performance of 19- and 75-year-olds.

Finally, the Age × Size interaction, F (3, 80) = 2.83, p < .05, was also significantly large. This effect was apparently produced by a curvilinear developmental trend with the 38.10-cm. test object, but a linear trend with the 7.62-cm. object. In particular, the 19-year-olds overestimated the width of the 38.10-cm.-wide form, but all other group means for the Age × Size factor represented underestimates. Thus, this interaction apparently resulted from an anomalously high mean by the 19-year-olds with the large standard object.

Neither the main effect of size nor any additional interactions involving size or Age × Distance were significantly large in

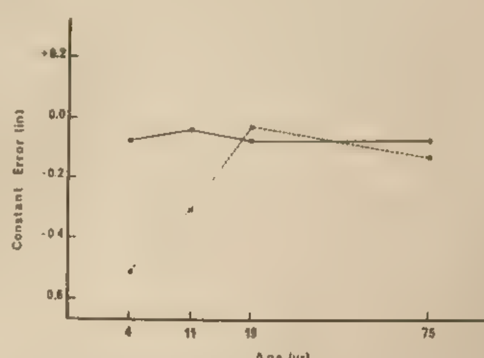


FIGURE 1. Mean constant errors (point of subjective equality minus standard) of judgments of the physical widths by *S*s who viewed the forms from 91.44-cm. (solid line) and 457.20-cm. (dashed line) distances. (Mean constant errors are averaged over sizes and angles of rotation.)

the analysis of these data. Apparently, the effect of presenting 7.62- or 38.10-cm.-wide standard objects at the 2 distances so that the same projective size is maintained does not decrease the effect of distance upon the development of shape constancy performance.

Discussion

The results of the present study show that shape constancy performance improves with age, at least between 4 and 19 yr. of age and with objects at a 457.20-cm. distance. The present results thus contradict those of Meneghini and Leibowitz (1967). As suggested by Gibson (1969, p. 365), it seems probable that the decreasing age trend found by Meneghini and Leibowitz resulted from the type of instructions used in their study. Their instructions can probably be termed "apparent" (Landauer, 1969). Such instructions have been shown to result in lower degrees of constancy performance and decreasing developmental trends (Kaess, 1970). When care is taken to insure *S*s' understanding of the task, shape constancy performance improves with age (Kaess, 1970, 1971). Klimpfinger's (1933) data also reveal an increasing developmental trend with children. However, the older groups of her study showed an apparent decrease of shape constancy performance, while the older groups of the present study do not. This discrepancy may also have resulted from differences in instructions, since Klimpfinger required projective shape judgments while the present study involved physical shape judgments. Apparently, when the perceptual task involves judgments of the physical shape of the object, shape constancy performance improves with age until adulthood and does not decrease with older *S*s.

Superimposed upon this developmental trend are the effects of other factors which affect the level of shape constancy performance. Shape constancy performance is more accurate when the object is nearly facing *S*, and when the object is at a 91.44-cm. distance. The reduction of shape constancy performance as a result of increased distance does not seem to be a product of the reduced visual angle subtended by the distant object, as expected, since there was no significant Age × Distance × Size interaction. Apparently, the "distance effect" is obtained with both large and small rectangles and is not entirely produced by differences in the visual angles of the standard and comparison forms.

It is possible that the distance effect results from the particular comparison technique of the method of limits paradigm used here. The present study involved variation of the distance of only the standard object. The comparison series was always presented at the near distance. Thus, the *S*s were required to match the standard stimulus to the comparison series forms when the distances of the forms were equal (near distance) and when the distances were disparate (far distance). A further test of the distance hypothesis in which the distance of the comparison series is varied simultaneously with that of the standard object would be of interest. In this manner it might be possible to determine whether the observational distance effect results from judgments of the standard form per se, or from changes in the ability to judge 2 forms, each of which is at a different distance.

Regardless of the locus of the effect, however, the significant Age × Distance interaction obtained in the present study suggests that the effect of observational distance depends upon the age of the *S*s. The results of the present study thus support the hypothesis of Vernon (1962) and Piaget (1967) that perceptual development proceeds from near to far space as age increases. At least with these experimental techniques, the effect of increased distance of the standard form is apparently greater with younger than with older *S*s. Consequently, developmental trends of shape constancy performance are similar to those for size constancy performance. Each task involves judgments of an "object-at-a-distance" (Gibson, 1969, p. 366), and developmental trends of performance on each task depend upon the observational distance involved.

References

Brault, H. Etude génétique de la constance des formes. *Psychologie Français*, 1962, **7**, 270–282.

Gibson, E. J. *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts, 1969.

Kaess, D. W. Form constancy and the perceptual task: A developmental study. *Journal of Experimental Psychology*, 1970, **83**, 465–471.

Kaess, D. W. A methodological study of form constancy development. *Journal of Experimental Child Psychology*, 1971, **12**, 27–34.

Klimpfinger, S. Die Entwicklung der Gestaltkonstanz vom Kind zum Erwachsenen. *Archiv für die gesamte Psychologie*, 1933, **88**, 599–628.

Landauer, A. A. Influence of instructions on judg-

ments of unfamiliar shapes. *Journal of Experimental Psychology*, 1969, **79**, 129–132.

MENEGHINI, K. A., & LEIBOWITZ, H. W. The effect of stimulus distance and age on shape constancy. *Journal of Experimental Psychology*, 1967, **74**, 241–248.

PIAGET, J. *The child's conception of space.* (Trans. by F. J. Langdon & J. L. Lunzer) New York: Norton, 1967.

VERNON, M. D. *The psychology of perception.* Baltimore: Penguin, 1962.

WOHLWILL, J. F. The development of "overconstancy" in space perception. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior.* Vol. 1. New York: Academic Press, 1963.

(Received January 15, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 22–26

ACQUISITION AND RETENTION OF MNEMONIC INFORMATION IN LONG-TERM MEMORY[1]

PAUL M. WORTMAN[2]
Northwestern University

PHILLIP B. SPARLING
Duke University

Paivio's (1969) 2-process theory of associative learning postulating differential effects on recall for word abstractness concreteness, and imaginal and verbal mediators, was evaluated within the context of mnemonic encoding. The mnemonic learning was self-paced, involving multiple lists with an immediate and delayed test of retention after 1 wk. The results supported a prediction of a trade-off in differential acquisition time for nondifferential recall performance in situations where presentation rate was not strictly controlled. A large amount of forgetting was found, independent of word or mediator type, thus supporting the notion that acquisition and retention are separate mechanisms. Finally, a highly significant effect for word type was found despite the relaxation in presentation rate, and its implication for modification of the theory was discussed.

The recent revival of interest in mnemonic techniques has led to a new theory of paired-associate (PA) learning (Bower, 1972; Paivio, 1969). According to this formulation, PA learning consists of 2 processes or mediational sets involving the formation of either verbal or imaginal associations. These processes interact differentially with 2 types of to-be-learned information—abstract and concrete words. Thus, image processes are supposed to be more predominant in learning concrete words than abstract words, since concrete nouns have readily available imaginal components of meaning, whereas abstract nouns do not. Verbal mediators, on the other hand, are assumed to be unaffected by changes in word type. While this predicted interaction has proved somewhat difficult to obtain, it has received empirical confirmation (Paivio & Foth, 1970).

The rather "elusive" nature of this phenomenon, however, raises a number of questions. On methodological grounds, how dependent is the interaction of word or item type and mediational set on the rather long study intervals used by Paivio and Foth (1970)? And, in particular, what happens when these intervals are made even longer? On the theoretical side, what is the relative importance of these interacting variables? And do they affect the retention of information in long-term memory (LTM)? The results reported by Paivio and Foth indicate a consistent superiority in learning of concrete over abstract PAs that is independent of encoding process, while the differential effects of these associative strategies seem highly dependent on temporal and instructional variables.

The present study seeks to assess this 2-process theory within the context of mnemonic learning. Mnemonic systems provide a convenient bridge between the theory and the study of LTM since they form the basis for the former and are also purported to enhance retention in the latter (Yates, 1966). Moreover, this interactive effect has not yet been observed with mnemonic strategies (cf. Paivio, 1968). In addition, this study examines mnemonic learning under circumstances typically employed with such techniques and ones which have not yet been subjected to empirical scrutiny. For example, nonlaboratory *S*s most often determine the rate at which items are learned, usually employ the mnemonic more than once, and expect to recall the information at some later time. Thus, the following experiment involved *S*-paced mnemonic learning of multiple lists

[1] This research was supported by U.S. Public Health Service Grant PE-00068-02 from the National Institutes of Health. The authors thank Gregory Kennington and Camille B. Wortman for their assistance in analyzing the data.

[2] Requests for reprints should be sent to Paul M. Wortman, Department of Psychology, Northwestern University, Evanston, Illinois 60201.

with both immediate and delayed tests of retention.

Under these conditions the 2 mediational processes should result in either differential time to master the to-be-learned information or differential retention of these items on an immediate test of retention. But, if Paivio (1969) is correct in assuming that image mediators differ in their availability, then both of these effects should not co-occur since differences in learning time should compensate for differences in immediate retention. Furthermore, if Olton (1969) is correct in that acquisition and retention (or forgetting) of information in LTM are separate processes, then these mediational sets should have no effect on delayed retention.

While mnemonic systems encompass a class of techniques, they are essentially similar in principle; one of these, in particular, has been employed in a number of experimental studies. This is the mnemonic rhyme or "peg-word" system, ONE-BUN, TWO-SHOE, ...TEN-HEN (Bugelski, 1968; Keppel & Zavortink, 1969; Paivio, 1968). This technique allows items to be learned by forming ordered pairs. For example, the first word would be memorized by associating it with the peg word BUN via a mediating image. The above studies have demonstrated that this mnemonic increases the rate of item acquisition and that imagery is essential for this to occur. Paivio developed a comparable rhyme composed of abstract words (i.e., ONE-FUN, etc.) but failed in his attempt to show its inferiority to the ONE-BUN mnemonic which is composed of high-imagery, concrete words. The following experiment seeks to evaluate the effects of these mnemonics on the acquisition and retention of information in LTM.

METHOD

Subjects and design. The *Ss* were 36 Duke University undergraduates whose participation fulfilled a course requirement. Each *S* was randomly assigned to one of 6 mnemonic acquisition conditions, so that there were 6 *Ss* per group in a $2 \times 3 \times 3 \times 2$ factorial design.

These various conditions represented 2 crossed, between-*Ss* variables, which included the abstractness-concreteness of the mnemonic rhyme pairs and mediational sets of bizarre images, common images, and verbal connections. While the 2-process theory of PA learning refers to the images usually or commonly available to and associated with concrete words, a number of studies have investigated the effects of bizarre images (Delin, 1968; Wood, 1967). Moreover, it is possible, especially under conditions where presentation time is not limited to a few seconds, that the formation of bizarre, as opposed to common, images reduces interference effects and thereby enhances recall. This would occur if such images were more elaborate and contained idiosyncratic meaning or if, as is likely, they were more differentiated and distinct than a common or ordinary image. For this reason, both of these visual imagery techniques were included along with the verbal mediation condition.

The last 2 factors in the design were composed of within-*Ss* variables. The first repeated measure included 3 lists of to-be-learned responses, and the second a temporal retention variable consisting of an immediate and delayed recall test.

Lists. Six 10-item lists of words were constructed using Paivio's tables (Paivio, Yuille, & Madigan, 1968). Three lists contained a total of 30 concrete words having a mean imagery (*I*) value (on a scale of 1–7) of 6.54 and a mean concreteness (*C*) value of 6.90. The 3 remaining lists were composed of 30 abstract words having a mean *I* value of 3.23 and a mean *C* value of 2.04. The *I* and *C* values for the 2 10-item peg-word mnemonics were approximately the same as those for the concrete and abstract word lists. All lists were counterbalanced for sequence of presentation within each 6-*S* group and for word frequency within each list.

Procedure. The procedure was patterned after Paivio and Foth's (1970) Experiment I. All *Ss* were run individually by the same *E*. Each *S* read a sheet containing a general outline of the experiment as well as providing the necessary cover story to avoid suspicion that they were returning a week later for a delayed retention test. The *Ss* then learned a number rhyme—the peg-word mnemonic. Half of them (18 *Ss*) were given the abstract rhyme—ONE-FUN, TWO-TRUE, THREE-FREE, FOUR-BORE, FIVE-LIVE, SIX-TRICKS, SEVEN-LEAVEN, EIGHT-FATE, NINE-KIND, TEN-MEND; the others learned the concrete rhyme—ONE-BUN, TWO-SHOE, THREE-TREE, FOUR-DOOR, FIVE-HIVE, SIX-STICKS, SEVEN-HEAVEN, EIGHT-GATE, NINE-WINE, and TEN-HEN. After hearing the mnemonic twice, all *Ss* were able to say it correctly aloud once and write it down twice. Each *S* was then given a sheet with appropriate illustrative examples explaining the use of the rhyme in learning word associations (by forming appropriate mediators), and containing 2 practice pairs to insure that he understood the instructions. This sequence was employed to avoid confounding learning the rhyme with using it to learn other words.

Each *S* then learned 30 words (3 10-word lists) employing one of 3 mediational methods to incorporate the list word and corresponding rhyme

word into either (*a*) a bizarre (strange or unusual) image, (*b*) a common (ordinary or usual) image, or (*c*) a sentence or phrase (verbal connection). For example, if the first word were GARDEN, the 3 mediators might be "flowers in the garden growing buns," "eating a bun in the garden," or "I saw a BUNny in the garden," respectively. Regardless of the mediational set, each *S* went through the rhyme list 3 times in learning the 30 words (i.e., one study trial for each list). The *S*s were cautioned to make each image or verbal connection distinct, thus hopefully preventing a linking together of 2 or 3 list words with the same corresponding rhyme word into one large image. To insure that *S* followed these instructions, he was required to write down (on another index card) either a description of the image mediator or the actual verbal connection. Paivio and Foth (1970) have reported that this method forces *S* to follow the particular mediational instructions.

All *S*s using the abstract rhyme learned abstract nouns and those using the concrete rhyme learned concrete nouns. The rationale for this was to avoid confounding the item attribute of the PAs with the mediational set and also to follow Paivio and Foth (1970), but with a between-*S*s mediation design. These authors noted that the interaction they found between these 2 variables may have resulted, in part, from their within-*S*s design.

Each *S* paced himself in learning the 30 words. He was given up to 1 min., if necessary, to form an appropriate image or phrase. This limit was seen as nonrestrictive since prior research (Bugelski, Kidd, & Segmen, 1968) has shown that 8 sec. is an optimal time to form an image. The *E* presented the words, one at a time, on index cards with one list following another.

Upon finishing the 3 lists, *S* was given a total recall test in which the numbers 1–10 were presented in random order, and he attempted to remember the 3 list words corresponding to the rhyme word for that number. Before ending the session, each *S* was then given 10 items from the Remote Associates Test (Mednick, 1962) to distract his attention and thus reduce rehearsal. One week later, the *S*s were given another total recall test during which they also wrote down the rhyme and all the mediators they had used to associate the peg words with the responses.

Results and Discussion

The results for the initial retention trial and their relation to the total time needed to form the associations to the rhyme words are depicted in Figure 1. Two *S*s—one in the concrete word, common image condition, and the other in the abstract word, bizarre image condition—failed to return for the delayed retention test, and their data were not included in the analysis. The mean numbers of words ini[illegible] re-called with both common an[illegible]rre imagery and verbal mediators, res[illegible]ely, were 23.8, 23.7, and 24.7 for concr[illegible]ns, and 15.2, 16.8, and 16.0 for abstra[illegible]ns. A univariate analysis of variance [illegible]med the usual highly significant main [illegible] for the item attribute of the mnemo[illegible]me pair (i.e., abstract or concrete), [illegible] 28) $= 25.33$, $p < .001$. As in Wood [illegible]peri-ments (1967), no differential effe[illegible] re-call was found for mediational se[illegible]e or interacting with the mnemonic rh[illegible] pair conditions (Fs < 1). However, [illegible] pre-dicted, the *S*s did take different a[illegible]ounts of time to acquire the same num[illegible]er of items, F (2, 28) $= 5.85$, $p < .01$. The bizarre images, not unexpectedly, took significantly longer to generate than either the common, F (1, 28) $= 5.18$, $p < .05$, or the verbal mediators, F (1, 28) $= 11.28$, $p < .01$, since they were not as readily available.

These findings appear to preclude differential overlearning resulting from self-pacing as a possible explanation. Where there were differences in pacing as in the mediation conditions, there was no difference in performance. But, this is precisely the situation in which such differences due to overlearning (i.e., having more study time) would be predicted. On the other hand, there was no difference in total learning time between abstract and concrete conditions ($F < 1.5$), but there was a large difference in retention. Similarly, the ability of all *S*s to learn the mnemonic rhyme perfectly on the first trial and to recall it correctly 1 wk. later seems to preclude differential overlearning of the rhyme and subsequent transfer effects as an explanation for these findings.

These results, moreover, clearly support the 2-process theory of verbal learning. There was a significant interaction in total learning or acquisition time between item attribute and mediational set for common images and verbal associations, F (1, 28) $= 5.33$, $p < .05$. The contrasts yielded the expected pattern with common images taking significantly longer to generate than verbal mediators for abstract nouns, F (1,

28) = 5.85, $p < .05$. On the other hand, the difference was not significant for concrete nouns ($F < 1$). Similarly, common images took less time to generate (i.e., they appear more available) for concrete than abstract words, F (1, 28) = 4.15, $p < .05$, while verbal connections were equally available for both ($F < 1.5$). The interaction, contrasts, and simple effects confirm both 2-process theory predictions of a trade-off between acquisition time and recall performance and the differential availability of imaginal mediators for concrete and abstract nouns.

On the delayed retention test 1 wk. later, the mean numbers of words recalled with both common and bizarre imagery and verbal mediators, respectively, were 16.2, 15.2, and 15.7 for concrete nouns, and 7.3, 8.0, and 8.2 for abstract nouns. A highly significant main effect was found for the retention variable, F (1, 28) = 118.73, $p < .0001$, indicating a large amount of forgetting during the week separating the 2 recall trials. In fact, this loss in retention was approximately constant for all conditions (about 8.2 items), showing no effects for either mediational set or item attribute (Fs < 1). Given the rather large initial differences in retention between abstract and concrete noun pairs, this constant forgetting resulted in a differential rate of forgetting (50% and 35%, respectively). However, this did not produce a significant interaction over the 2 retention trials ($F < 1$).

This supports Olton's results (1969) indicating that acquisition and retention in LTM are separate processes with mnemonic encoding affecting the former, but not the latter, process. Moreover, this constant forgetting contradicts Bower's (1972) notion that the decay rate in LTM is faster for verbal mediators than for imaginal ones. Instead, it necessitates a clear distinction between availability and decay. The former affects acquisition and the latter retention; and as these data indicate, it is mediator availability in mnemonic learning that reflects the interaction of associative techniques and item type. Moreover, the similarity and consistency of the between-Ss findings for item attribute

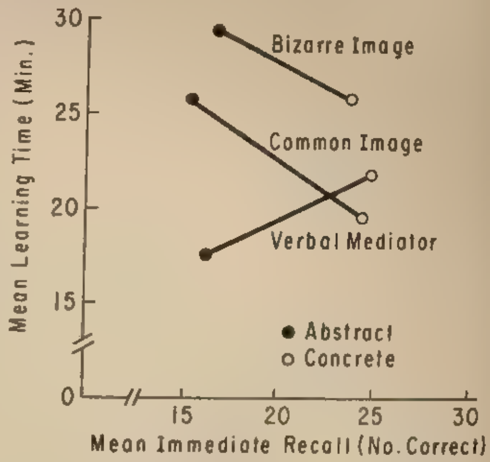


FIGURE 1. Mean immediate recall and learning time as a function of item concreteness abstractness and mediational technique.

and mediational set over the 2 retention tests indicate the absence of differential effects on the second trial resulting from the first retention probe.

There is, in addition, no differential or learning-to-learn effect for the 3 lists ($F < 1.5$). It is possible, however, that the amount of forgetting from LTM was overestimated due to a procedural error. Since the first recall test followed immediately after the study trial for the third list, with no interpolated activity, not all of the items recalled may have been in LTM. The mean amounts of forgetting from the 3 lists were 2.2, 2.5, and 3.6 items, respectively. This produced a significant List × Retention Trial interaction, F (2, 27) = 9.17, $p < .01$, with forgetting from List 3 significantly greater than that from the other 2 lists, both Fs (1, 28) > 6, $p < .05$. Thus retention was underestimated by approximately one item—a relatively small effect compared with the large number forgotten. Finally, while mediator availability has been found to be an important factor in the acquisition or encoding of an item, it is also possible that mediator decodability affects performance. An analysis of the missed items on the delayed recall trial revealed that it was more difficult to decode the responses from abstract mediators (a mean of 3.0 items not de-

coded) than concrete mediators (a mean of 1.2 undecoded items), $F(1, 28) = 7.53$, $p < .05$.

Conclusions

The results of this study indicate that when (mnemonic) learning is S paced instead of E paced, there is a trade-off in the performance for acquisition or learning time. That is, S-paced learning results in differential rates of acquisition for various mediating strategies but not in differential performance, while the reverse is true in E-paced studies (where the presentation rate is typically less than 15 sec.). The availability hypothesis that common images are more readily evoked by concrete than abstract nouns, with no such difference occurring with verbal mediators, was also supported. While these findings are in agreement with Paivio's (1969) 2-process theory of verbal learning, the relatively large effects on recall due to abstractness–concreteness, despite the S-paced differences in acquisition rate, indicate that a modification of theory is necessary. The factor of item attribute absolutely affects performance independent of presentation rate, and as such is prepotent; while mediational set primarily affects the information-processing rate (i.e., item acquisition in LTM) and only incidentally (i.e., at fast presentation rates) the performance of the learning system.

REFERENCES

Bower, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1972.

Bugelski, B. R. Images as mediators in one-trial paired-associate learning. *Journal of Experimental Psychology*, 1968, **77**, 328-33[illegible]

Bugelski, B. R., Kidd, E., & Segmen, J. Image as a mediator in one-trial paired-associate learning. *Journal of Experimental Psychology*, 1968, **76**, 69 73.

Dilin, P. S. Success in recall as a function of success in implementation of mnemonic instructions. *Psychonomic Science*, 1968, **12**, 1[illegible] 154.

Keppel, G., & Zavortink, B. Further test of the use of images as mediators. *Journal of Experimental Psychology*, 1969, **82**, 190–19[illegible]

Mednick, S. A. The associative basis of the creative process. *Psychological Review*, 1962, **69**, [illegible]0-232.

Olton, R. M. The effect of a mnemonic upon the retention of paired-associate verbal material. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 43 48.

Paivio, A. Effects of imagery instructions and concreteness of memory pegs in a mnemonic system. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, [illegible]77–78.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241 -263.

Paivio, A., & Foth, D. Imaginal and verbal mediators and noun concreteness in paired-associate learning: The elusive interaction. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 384 390.

Paivio, A., Yuille, J. C., & Madigan, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Pt. 2).

Wood, G. Mnemonic systems in recall. *Journal of Educational Psychology Monograph*, 1967, **58**(6, Pt. 2).

Yates, F. A. *The art of memory.* London: Routledge & Kegan Paul, 1966.

(Received January 31, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 27 36

A RATIO RULE FROM INTEGRATION THEORY APPLIED TO INFERENCE JUDGMENTS[1]

MANUEL LEON[2] AND NORMAN H. ANDERSON

University of California, San Diego

The *S*s judged the likelihood that a sample had come from 1 of 2 binary urns of specified proportions of red and white beads. The compositions of the sample and of the 2 urns were varied in a 3-way design. A ratio model from integration theory fit the data quite well. The response is treated as a resultant of 2 competing response tendencies that reflect the felt likelihoods that the sample comes from either urn. A novel application of analysis of variance for nonlinear models was employed in the test of fit. Relations to Bayesian theory, choice theory, representativeness, as well as previous work on serial integration by J. C. Shanteau and T. S. Wallsten, were discussed.

This report considers the 2-urn task that has been used in much of the experimental work on the Bayesian approach to decision making. A single sample was drawn from 1 of 2 specified binary urns, of 80/20 and 40/60 composition, for instance, and *S* judged the probability that the sample had come from the 80/20 urn. However, instead of applying the Bayesian statistical model (Edwards, 1968; Slovic & Lichtenstein, 1971), the present report follows the lead of Shanteau (1970, 1972) in seeking a descriptive, psychological model for subjective inference.

In preliminary work, conversation with *S*s suggested that their judgments were based on a comparison between the 2 probabilities that the sample came from either urn. On that basis, the overt response is considered as a compromise between 2 covert response tendencies. Upon seeing the sample, *S* evaluates the subjective probabilities that it comes from either urn. These 2 response tendencies then determine the actual response. The specific question at issue is the integration rule that governs how the 2 response tendencies combined.

Two particular integration rules suggested themselves. One was a subtracting model from conflict theory (Anderson, 1962b). The 2 competing response tendencies would be conceptualized as approach and avoidance tendencies; the effective response would equal the difference in their strengths.

The other integration rule derives from averaging theory (Anderson, 1974). Each covert response tendency would be assumed to have a scale value and a weight, e.g., s_1 and w_1 for the more likely urn, s_2 and w_2 for the less likely urn. The response is then the weighted average:

$$R = (w_1s_1 + w_2s_2)/(w_1 + w_2). \quad [1]$$

In the present task, the scale values may be specified as $s_1 = 1$, $s_2 = 0$, since these represent the end points of the decision scale. Equation 1 then reduces to

$$R = w_1/(w_1 + w_2). \quad [2]$$

Equation 2 is a ratio rule: The overt response is the ratio of the strength of 1 response divided by the sum of the strengths of the 2 responses. This ratio rule can be generalized directly to the case of several competing responses, corresponding to several possible parent urns. In this situation, therefore, integration theory leads to a ratio rule similar to that used by Luce (1959) and others. Unlike Luce's choice theory, however, integration theory

[1] This research was supported by National Science Foundation Grant GB-21028 and facilitated by National Institute of Mental Health grants to the Center for Human Information Processing, University of California, San Diego. We wish to thank E. F. Alf, M. H. Birnbaum, G. C. Oden, J. C. Shanteau, and T. S. Wallsten for their helpful comments on an earlier draft.

[2] Requests for reprints should be sent to Manuel Leon, Department of Psychology, University of California, La Jolla, California 92037.

allows deterministic numerical response measures.

In testing such algebraic models, functional measurement methodology typically employs factorial designs. An attempt was made in the present experiment to avoid any assumptions about relations among samples of different composition. Accordingly, single samples were used, and the factors of the design were the compositions of the more likely urn and of the less likely urn, a device also used by Wallsten (1972). The subtracting model then predicts a zero between-urns interaction in the analysis of variance (Anderson, 1962b). The ratio model is nonlinear and not testable so simply. However, an important new test procedure was developed that may be generally useful in the analysis of nonlinear models.

Method

On each trial, *S* saw a single sample of red and white beads and 2 binary parent urns. His task was to choose the more probable urn and then make a graphic rating "to indicate on the scale how likely it is that you are correct."

Design

The main experimental design was a 3 × 4 × 8 factorial. There were 3 levels of the *less probable urn* in which the red/white ratios were 10/90, 20/80, and 40/60. There were 4 levels of the *more probable urn* in which the red/white ratios were 60/40, 70/30, 80/20, and 90/10. Finally, there were 8 *samples:* 4 homogeneous, with 1, 2, 3, or 4 red beads, and 4 mixed, with 3/1, 4/1, 3/2, and 6/2 red/white ratios. This design was balanced over red and white to yield 192 judgments. In addition, 16 extreme anchor sets were included.

An auxiliary design was also employed in which the parent urns were 70/30 and 30/70 for all samples. There were 100 samples, which included all possible red/white ratios in which the number of red and white beads each varied from 1 to 10. Four additional extreme anchor sets were also used in this condition.

Procedure

Each urn was represented by a transparent plastic bag with 100 beads in the specified proportion, together with a numerical label, such as 60/40, to indicate its contents. On each trial, an index card was presented on which the 2 relevant urns were listed (e.g., 60/40 and 20/80), and on which the sample was represented by rows of large red and white dots, one color in each row.

Responses were made by placing a pin in a [illegible]-cm. scale, read from the rear by E. The right and left halves of the scale were used to indicate likelihood judgments of the mostly red and the mostly white urns, respectively, with the center of the scale corresponding to equally likely.

The judgments were collected in the following 4 consecutive stages: (*a*) preliminary instructions followed by a practice phase which included the 16 anchor sets plus 64 sets from the main design, (*b*) the 16 anchors plus the 192 sets from the main design, (*c*) the 104 judgments of the auxiliary design, and (*d*) direct judgments of the likelihood of each of the 8 samples coming from each 1 of the [illegible] urns used in the main design. These last data did not seem very useful and are not reported here. Except for practice, the deck of index cards was shuffled separately for each *S* for each stage. Stimulus cards were presented at approximately one every [illegible] sec.

The *S*s were 24 introductory psychology students who served as part of a course requirement. Each served in a single session of approximately [illegible] min. with a 5-min. rest after the main design.

Results

Empirical Results

Main design. Each curve in Figure 1 shows the mean response to 1 of the 8 samples, plotted as a function of the more probable urn. The 2 main trends are as expected: the likelihood judgment increases (*a*) as the composition of the more probable urn becomes more extreme, and (*b*) as the difference between red and white beads in the sample becomes more extreme. The main effects of these 2 variables were highly significant, *F*s $(3, 69) = 72.55$ and 36.92, as was also the effect of the less probable urn, $F\ (2, 46) = 58.11$.

Neither the difference between red and white, nor the ratio of red to white, is the sole determinant of judged likelihood. First, to each homogeneous sample in the left panel there is a corresponding mixed sample in the right panel that has the same difference between red and white. Since the 2 sets of curves have different overall shape and elevation, the response is determined by more than the difference between red and white. Second, the curve for the 6/2 sample is substantially higher than for the 3/1 sample. Thus, the response is determined by more than the sample ratio. Supporting evidence is given in the next subsection.

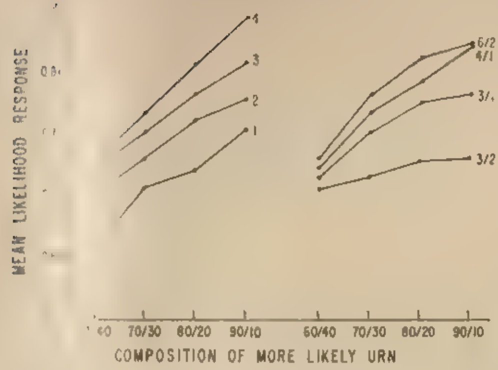


FIGURE 1. Mean likelihood judgment as a function of the composition of the more likely urn, and of sample composition as listed by each curve. (Data averaged over less likely urn.)

Special theoretical significance attaches to the case in which the 2 urns have symmetrical red/white proportions. From the Bayesian model, the statistically optimal response then depends only on the difference between red and white. Descriptive models might also embody such an assumption on the ground that each red–white pair of beads is uninformative and would be ignored. Because of the theoretical significance of these data, they are presented separately in Figure 2.

The relevant comparison in Figure 2 is between the left and right panels. They should be identical if the response is determined by the difference between red and white. But in fact the 2 panels differ considerably. The response to homogeneous samples is more extreme in either direction, and covers a greater vertical range. Also, the mixed samples show no effect of set size for the 60/40 urn pair. That the 2 panels are reliably different is shown by the Sample Type × Urn Composition interaction, F (2, 46) = 6.65.

Auxiliary design. In this design, the 2 urns were complementary, of 70/30 and 30/70 composition. All 10 × 10 possible samples, with 1–10 beads of each color, were used. Illustrative data are shown in Figure 3.

In the left panel of Figure 3, the near identity of the 3 curves shows that the response was completely governed by the difference between the number of red and white beads in the sample (plotted on the horizontal). At any point, the curves represent samples that have the same difference between red and white, though differing in sample size. The curve labeled 9 + 10 is averaged over the samples in which there were 9 or 10 beads of the dominant color. The curves labeled 7 + 8 and 5 + 6 were obtained similarly. It was not feasible to plot all the data points because they all lay on top of one another. As these curves show, however, sample size per se had no effect in this phase of the experiment.

The same conclusion is illustrated in different form in the right panel of Figure 3. The solid curve represents the 5 samples with a 1/2 ratio. The dashed curve represents the 3 samples with a 2/3 ratio. The upward trend shows that something more than the sample ratio controls the response. The near equality of the 2 curves illustrates again that the sample difference is the controlling factor.

The data from this auxiliary design should be interpreted cautiously. The very simplicity in the pattern of results suggests that Ss may have adopted a cancellation strategy. Canceling red against white would lead to the obtained results. Since the 2 urns were complementary and fixed over all trials, Ss could devote their attention to the characteristics of the sample itself.

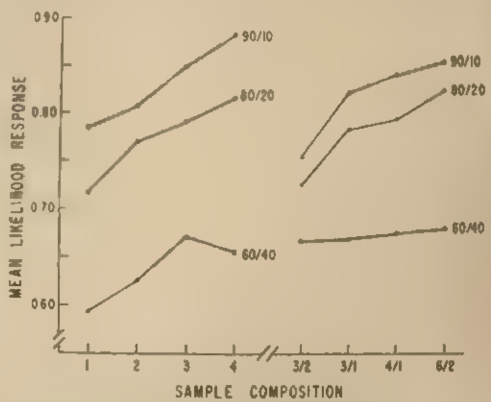


FIGURE 2. Mean likelihood judgment as a function of sample composition and urn composition. (Data for symmetrical urns only.)

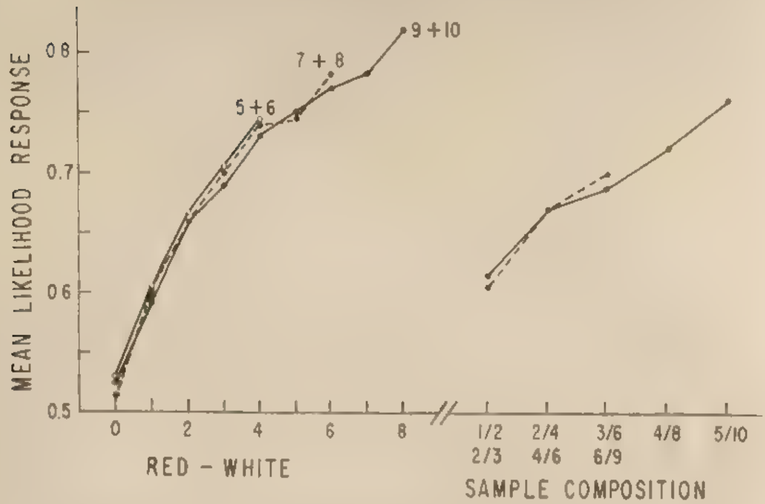


FIGURE 3. Mean likelihood judgment from the auxiliary design with 70/30 and 30/70 urns.

In the main design, such a cancellation strategy would seem less appropriate. In fact, the difference between the left and right panels of Figure 1 shows that something more than the sample difference controlled the response. That was true even for the cases in which the urns were symmetric as already noted. Deviations from the difference rule were moderately small, however, so it may be a good first approximation.

Model Analysis

Model predictions. Figure 4 plots mean likelihood response as a function of the more and less likely urns. The left and right panels are averages over the 4 homogeneous samples and the 4 mixed samples, respectively.

The curves in Figure 4 are theoretical, obtained by fitting the ratio model of Equation 2 as described below. The fit is quite close which supports the ratio model.

An additional aspect of Figure 4 is the nonparallelism of the data points in each panel. A model that assumed 2 response tendencies that combined by a subtracting rule would predict parallelism. That the observed nonparallelism is reliable is shown by the significant More Likely Urn × Less Likely Urn interaction, $F(6, 138) = 2.99$. Wallsten's (1972) data, converted from odds to probabilities, show the same shape, with an even more marked interaction. On the face of it, therefore, a subtracting model does not account for the data.

Figure 4, incidentally, illustrates an optical illusion that affects apparent parallelism, as the data seem more parallel than they are in fact. The vertical spread of data points for the 60/40 urn is greater than for the 90/10 urn by about 75% for the homogeneous samples and 30% for the mixed samples. In making graphic assessments of parallelism, it is often helpful to turn the graph sidewise and measure the physical distances.

Weight estimates. Figure 5 plots mean weight estimates from the ratio model as a function of urn and sample composition. The trends can be understood most easily by reference to the theory, in which the values of w_1 and w_2 represent the response strengths of the 2 alternatives.

Theoretically, therefore, w_1 should increase as the composition of the more likely urn becomes more extreme, and also as sample composition becomes more extreme. Both trends are quite marked.

The trends for w_2, in contrast, are much less marked. A mild effect of the composition of the less likely urn can be seen, but

the effect of the sample was so small that the points could not be graphed separately. To a first approximation, therefore, the less likely urn appears to have a nearly constant response strength.

Estimation and goodness of fit. The model was tested by a novel and potentially important application of functional measurement methodology. Since Equation 2 is nonlinear, the parameter estimates are not statistically independent, and the number of degrees of freedom used in the estimation is unknown. If only relatively few parameters were estimated, the rule of one degree of freedom per parameter would presumably be an adequate approximation. Even so, there would be a difficult problem of incorporating the variabilities of both the observed and the predicted data points in the test. In general, exact analysis of nonlinear models is not easy.

The method developed here may bypass all these difficulties in a simple way. Chandler's (1969) STEPIT routine was applied to estimate the weight parameters of Equation 2 for each of the 8 samples for each *S*. The means of these estimates were used in Figure 5. Predicted responses were also generated for each of the 8 samples for each *S* by using these weight estimates in Equation 2. The means of these predicted responses were used in Figure 4.

To test goodness of fit, the predicted responses were subtracted from the observed to yield deviation scores. The data for statistical analysis are thus the 8 3 × 4 matrices of deviation scores for each *S*.

These deviation scores were analyzed in an ordinary repeated-measurements analysis of variance. The homogeneous and mixed samples were treated separately so that there were 2 3-way analyses of variance. If the model is correct, then these deviations should be no more than chance. The *F* ratios in Table 1 should cluster around 1 and none should be significant. In fact, 1 of the 16 is significant, but that is

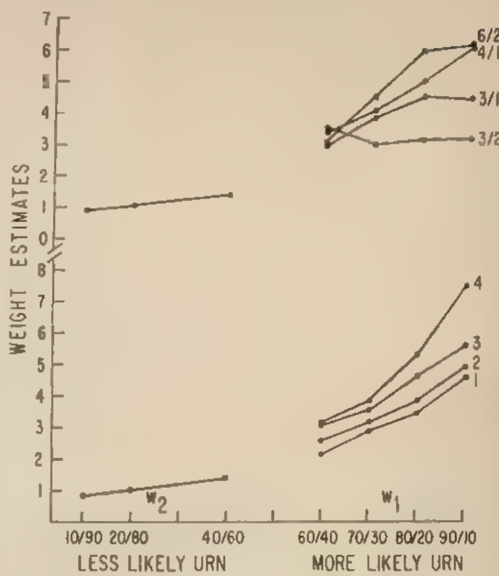


FIGURE 5. Response strengths of more and less likely choice alternatives.

TABLE 1

SUMMARY ANALYSES OF VARIANCE OF DEVIATIONS FROM RATIO MODEL

Source	*df*	Homogeneous samples		Mixed samples	
		F	Error	*F*	Error
Mean	1	1.36	.2568	1.22	.1169
Samples (S)	3	.45	.1294	1.78	.1069
Less likely urn (LLU)	2	1.60	.7615	1.13	.4058
More likely urn (MLU)	3	.82	.7956	1.00	.5465
S × LLU	6	1.21	.4683	1.63	.3571
S × MLU	9	1.02	.3685	.95	.3741
LLU × MLU	6	.75	28.2653	.61	21.1473
S × LLU × MLU	18	1.15	18.3115	2.01*	17.9208

Note. Error mean squares have 23 times as many degrees of freedom as corresponding systematic source. Error mean squares multiplied by 10,000.
* $p < .05$.

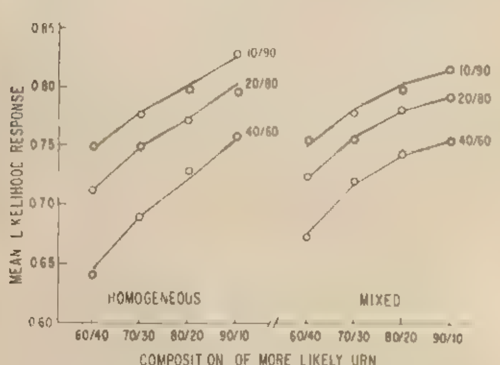


FIGURE 4. Predicted and obtained likelihood judgments. (Curves are theoretical, from ratio model. Data points averaged over 4 samples in each panel.)

about the chance expectation with 16 tests. Thus, this goodness of fit test supports the model.

This test may be disturbing because a good fit is guaranteed by using 7 parameters to fit 12 data points. However, the error term is itself composed of the small deviations from that good fit. Thus, the test should be sensitive even to small discrepancies if they are systematic across *S*s.

The main technical problem in this test is that the predicted scores, and hence also the deviation scores, are not independent but are correlated through their joint dependence on the estimated parameters. However, the analysis of variance for repeated measurements allows for correlated scores (Winer, 1971, p. 279).

Two comments may help relate this test to the ordinary applications of analysis of variance. First, any linear contrast on the deviation scores would have an exact t test, obtained by comparing its mean to its variability across *S*s. Applied to main effects or interactions on one degree of freedom, this t test would be equivalent to the above F test. Second, if the present method were applied to an additive rather than a nonlinear model, it would be equivalent to the ordinary analysis of variance except that all main effects would necessarily be zero.

Tests based on group data are sensitive to discrepancies that are systematic across *S*s. However, real discrepancies at the individual level may get overlooked if they are idiosyncratic. Table 1 contains a warning in this regard since the error mean squares for the last 2 interactions are much larger than for the others. Since the error term for each listed source is its interaction with *S*s, the size of these errors suggests that there may be real discrepancies though of different form for different individuals.

The present method could be applied directly to single *S*s by replicating the experiment several times. These replications would be treated exactly as *S*s in the present analysis. The deviation scores within each replication would be correlated, but such correlation is allowed in the given analysis.

This method for testing goodness of fit is presented tentatively. If it is supported in further work, it could be extremely important because it provides an exact analysis for nonlinear models. An especially interesting possible application is to response scaling. Thus, the test could be used to provide an error theory for nonmetric approaches, such as used in conjoint measurement (Krantz, Luce, Suppes, & Tversky, 1971), or Kruskal's (1965) monotone analysis of variance, for which an error theory is presently lacking (see Weiss & Anderson, 1972). In such applications, the test would be made on the differences between the monotonically transformed data points and the best-fit additive representation.

Discussion

Ratio Model

The basic assumption of the present approach is that the overt response is a compromise between 2 competing response tendencies. One specific integration rule considered here was the ratio rule from averaging theory, i.e., that the overt response was the average of the 2 competing response tendencies. This ratio rule gave a good account of the data and led to meaningful parameter estimates. The strength parameter for the less likely urn was approximately constant over conditions, as though it exerted much the same effect regardless of its composition or the composition of the sample. In contrast, the strength parameter of the more likely urn varied markedly across conditions, suggesting that this urn was given the main attention.

An alternative, the subtracting model, was also considered in which the overt response is just the difference between the 2 response strengths. Interestingly enough, the ratio and subtracting models cannot be distinguished in the present experiment as they are monotonically equivalent. If R obeys a ratio model, then $\log(R^{-1} - 1)$ obeys a subtracting model. Since functional measurement methodology allows for monotone response transformation (Anderson, 1962b), the present data do not allow the conclusion that the ratio model is superior to the subtracting model. Either might be correct.

Some preference may be expressed for the ratio model because it fits the raw data,

whereas response transformation would be required for the subtracting model. Previous work on functional measurement has shown that raw ratings are often near-interval scales and that argues for the tentative acceptance of the ratio model. On the other hand, parameter estimates for the 2 urns would be more comparable under the subtracting model.

The ratio model and the subtracting model are not monotonically equivalent in general. For 3 or more urns, the ratio model is inherently nonlinear in the sense that no monotone transformation will make $w_1/(w_1 + w_2 + w_3)$ linear in all 3 variables at once. The subtracting model, $w_1 - w_2 - w_3$, remains linear, at least in form.

A limitation on the applicability of at least the ratio model is suggested by the marked difference between the 2 weight parameters obtained here. Whereas the weight parameter for the more likely urn showed a strong dependence on the stimulus variables, the weight parameter for the less likely urn was almost constant. This suggests that the parameter valuation process may be configural, influenced by attentional factors. To ensure constant parameters, it would thus seem advisable to keep the same urn predominant in all cases.

Choice Theory

Although the ratio model of integration theory has the same form as Luce's (1959) choice model, there are several basic differences between them. The most important of these center on the difference between choice probabilities and numerical response measures. Luce's model applies, in effect, only to imperfect discrimination, or probabilistic threshold choices. In contrast, integration theory deals with numerical-valued response strengths, and applies even when there is no probabilistic element in the choice.

This difference is well illustrated in the present task, in which the appropriate response is never in doubt. Unless the sample is split evenly, the appropriate choice of urn is specified by the majority proportion in the sample. Since the choice probabilities are 0 or 1, Luce's formulation does not apply.

Whether the present averaging model might be applicable to threshold choices is uncertain. Substantively, if not formally, that would seem to require that the actual choice process can be conceptualized as an averaging process even in the case of imperfect discrimination. Some evidence that supports this idea is given by Friedman, Carterette, and Anderson (1968), Friedman, Rollins, and Padilla (1968), Himmelfarb (1970), Robbins and Medin (1971), and Levin, Dulberg, Dooley, and Hinrichs (1972).

Bayesian Theory

The Bayesian approach is similar to integration theory in its emphasis on problems of combining information to reach a judgment. The wide use of the 2-urn task reflects the pioneering efforts of the Bayesian workers in studying how information is integrated in decision making.

In a strict sense, comparison of these 2 approaches is not appropriate. Integration theory is a descriptive psychological theory, whereas the normative Bayesian model derives from mathematical statistics and prescribes the optimal response in an actuarial sense. In various practical applications, knowledge of the optimal, normative response can be most important.

Almost inevitably, however, normative models are forced into a descriptive mode for which they are ill-suited. Thus, a dominant concern in the Bayesian work has been with explanations of "conservatism," the general tendency for the response to be less extreme than the optimal response. Conservatism was obtained in the present experiment also, but no attempt has been made to explain it because there is nothing to explain. Conservatism is not a phenomenon, only a name. It exists only by reference to a statistical model that has no psychological relevance. "Explanations" of conservatism in terms of misperception, misaggregation or response bias are inappropriate because there is no psychological phenomenon to be explained.

To be fair to the Bayesians, it should be repeated that conservatism can be important in practical applications in which there is a cash value on correct responding. Pursuing such problems in the experimental laboratory can be quite useful. Nevertheless, commitment to the study of optimal behavior is not innocuous. It has diverted attention from the actual behavior, as is implied by the papers of Shanteau (1970, 1972), Slovic and Lichtenstein (1971), Wallsten (1972), and Anderson (1971, p. 195). In particular, the data reported in the Bayesian studies are ordinarily derived statistics, calculated from the Bayesian model itself, so that identical behavior in different conditions will produce different derived

statistics. Such data analysis makes it hard to understand what *S*s actually did.

Representativeness

Kahneman and Tversky (1972) have argued from a heuristic of sample-population "representativeness" that the response to single samples in the Bayesian task should be controlled almost entirely by the sample proportion. In particular, they claim that the response should be virtually independent of the sample size, and independent of the urn proportions (pp. 446, 450). The data of their Table 1 support their claim since the responses are very nearly equal to the actual sample proportions.

The present results differ radically from those of Kahneman and Tversky (1972). Figures 1, 2, and 4 show a very strong effect of the urn proportions, in agreement with the results of Shanteau (1972) and Wallsten (1972). Further, the right panels of Figures 1 and 3 show a strong effect of sample size with sample proportion held constant. In both respects, therefore, the present data disagree with those given by Kahneman and Tversky.

The simplest explanation of this difference is that Kahneman and Tversky (1972) failed to make their task clear. Their *S*s were run in large classroom groups in 1 or 2 min. of a regular class session, and each *S* judged only one problem. That the task was not clear is suggested by the fact that about 10% of the *S*s had to be eliminated (pp. 432, 447). It hardly seems surprising, therefore, that their *S*s might have failed to understand the task and simply interpreted their job as one of judging the sample proportion.

A warning on this problem was sounded by Shanteau (1970) who found no difference between estimation and inference instructions with serial presentation. Logically, these 2 instructions should produce quite different behavior, with the inference group tending toward 0 or 1, and the estimation group toward the prevailing proportion. The lack of difference was surprising, therefore, yet it was replicated in a follow-up experiment in which Shanteau used special care in explaining the judgment task to *S*s.

Shanteau's (1970) results indicate that *S*s in the Bayesian task are not making inference judgments, as has been assumed in tests of the Bayesian model. Instead, they appear to be doing something qualitatively different, more akin to estimation. This illustrates an important methodological point that appears to be especially relevant to judgment theory. The meaning of a response is not given simply by the words that the investigator uses, but rather by the network of empirical relations that surround the behavior.

Simultaneous Integration

The present work is most closely related to that of Shanteau (1970, 1972) and of Wallsten (1972). In all this work, primary emphasis is on the algebraic rule or psychological law that governs the integration of the information. Wallsten attempted to develop a descriptive analogue of the Bayesian model, whereas Shanteau employed a model derived from information-integration theory. Shanteau was mainly concerned with serial integration which is discussed below. Simultaneous presentation, as used here, seems to involve different processes and requires a few preliminary comments.

A basic question for simultaneous presentation is whether the effect of the set of stimulus information can be represented as a simple function of the elements of the set. Considerable evidence for this view has been obtained in person perception and various other tasks (e.g., Anderson, 1962a, 1967, 1974).

In the 2-urn task, a similar view has been taken by Shanteau (1970, Equation 3) and by Wallsten (1972). In particular, Wallsten's multiplying assumption implies that the difference between the number of red and white beads is the only relevant property of the sample for the special case of symmetric urns. This follows from Wallsten's Equation 3a, together with the added assumption that a sample split evenly between red and white has no effect when the urns are symmetric.

Unfortunately for the 2-urn task, the present data indicate that the effect of the sample is not a simple function of its elements. The qualitative test of Figure 2 shows that the difference between red and white is not the sole determiner of the response. Sample composition appears to have a gestalt effect. The data from the auxiliary design (Figure 3) did obey a cancellation property, but that may reflect a more deliberate strategy potentiated by special task conditions. Shanteau's (1970, Figure 4) data also imply that the effect of a simultaneous sample is not a simple function of its elements. This problem, it should be noted, did not arise in Shanteau's results with serial presentation, but he used only one-bead samples.

Serial Integration

Two basic reports by Shanteau (1970, 1972) have studied behavior in the Bayesian task using sequences of one-bead samples. A serial integration model gave a very good account of the data whereas the Bayesian model did poorly. A surprising result, noted above, was the lack of difference between estimation and inference instructions. This result caused a serious conceptual problem for the Bayesian formulation because the evidence suggested that *Ss* were doing something akin to estimation regardless of the instructions.

However, Shanteau's (1970, 1972) success with the observed behavior obscured a theoretical difficulty with his model as applied to inference rather than estimation. No problem arises if the same sample is presented at each serial position; unless it was split evenly, the response would always asymptote at 0 or 1. However, that would not necessarily happen if sample composition varied across serial position. In particular, a sequence of random samples of single beads from any urn would asymptote at the urn proportion. Despite its success, therefore, the serial integration model used by Shanteau would not seem able to handle general inference judgments without further elaboration.

Wallsten (1972) has also attempted to develop a model for serial integration. However, it has been worked out explicitly only for the case of a sequence of identical samples, and in that form it is effectively equivalent to the serial integration model used by Shanteau. Wallsten's approach seems more general since it explicitly allows for monotone response transformation, but the functional measurement methodology used by Shanteau has the same capability. Since Shanteau's analysis showed that the untransformed response was already an interval scale, response transformation was not needed. Wallsten's nonmetric, ordinal analyses did not allow a satisfactory test of the basic model.

The present formulation deals explicitly only with a single sample, and it is uncertain how it should be extended to handle serial integration of successive samples. Three possibilities deserve mention. The first is that separate cumulative counts of the red and white beads are maintained, and that the 2 competing response tendencies are recomputed from the cumulative counts after each successive sample. In effect, this reduces the serial case to the simultaneous case. However, it would allow serial position weights for each separate count so that serial position effects could be allowed for.

The second possibility is similar to the first, but allows only one counter which cumulates the difference between red and white. Weighting as a function of serial position would be allowable, but a cancellation strategy would be required within any single sample. Cancellation is not entirely consistent with present results, though it might provide a reasonable first approximation.

The third possibility is that each later sample sets up 2 competing response tendencies in exactly the same way as with the first sample. These are then averaged in with the single resultant response produced at the previous serial position. The resultant would have a scale value of 0 or 1, and a weight that would reflect the total amount of information.

Concluding Comments

Three theoretical problems are basic in the analysis of the 2-urn task. The first problem concerns the integration model for a single sample. Present data give some slight preference for the ratio model over the subtracting model, but a critical test between them remains to be done.

The second problem is how the response to any sample relates to its component elements. Unfortunately, as noted above, this problem may not have a simple solution like that found for verbal stimuli. Present data indicate some departure from the simple cancellation strategy even with symmetrical urns.

The third problem is that an adequate general theory must handle serial presentation as well as simultaneous presentation. Serial presentation involves additional complications, however, and it may be desirable to place greater attention on the first 2 problems which have been relatively neglected in previous work.

REFERENCES

Anderson, N. H. Application of an additive model to impression formation. *Science*, 1962, **138**, 817-818. (a)

Anderson, N. H. On the quantification of Miller's conflict theory. *Psychological Review*, 1962, **69**, 400-414. (b)

Anderson, N. H. Averaging model analysis of set size effect in impression formation. *Journal of Experimental Psychology*, 1967, **75**, 158-165.

ANDERSON, N. H. Integration theory and attitude change. *Psychological Review*, 1971, **78**, 171–206.

ANDERSON, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*. Vol. 2. San Francisco: Freeman, 1974.

CHANDLER, J. P. Subroutine STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, **14**, 81–82.

EDWARDS, W. Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.

FRIEDMAN, M. P., CARTERETTE, E. C., & ANDERSON, N. H. Long-term probability learning with a random schedule of reinforcement. *Journal of Experimental Psychology*, 1968, **78**, 442–455.

FRIEDMAN, M. P., ROLLINS, H., & PADILLA, G. The role of cue validity in stimulus compounding. *Journal of Mathematical Psychology*, 1968, **5**, 300–310.

HIMMELFARB, S. Effects of cue validity differences in weighting information. *Journal of Mathematical Psychology*, 1970, **7**, 531–539.

KAHNEMAN, D., & TVERSKY, A. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 1972, **3**, 430–454.

KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. *Foundations of measurement*. Vol. 1. New York: Academic Press, 1971.

KRUSKAL, J. B. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society* (Series B), 1965, **27**, 251–263.

LEVIN, I. P., DULBERG, C. S., DOOLEY, [illegible] F., & HINRICHS, J. V. Sequential dependencies in single-item and multiple-item probability learning. *Journal of Experimental Psychology*, 1972, **93**, 262–267.

LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.

ROBBINS, D., & MEDIN, D. L. Cue selection after multiple-cue probability learning. *Journal of Experimental Psychology*, 1971, **91**, [illegible].

SHANTEAU, J. C. An additive model for sequential decision making. *Journal of Experimental Psychology*, 1970, **85**, 181–191.

SHANTEAU, J. C. Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology*, 1972, **93**, [illegible]–68.

SLOVIC, P., & LICHTENSTEIN, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, **6**, 649–744.

WALLSTEN, T. S. Conjoint-measurement framework for the study of probabilistic information processing. *Psychological Review*, 1972, **79**, 245–260.

WEISS, D. J., & ANDERSON, N. H. Use of rank order data in functional measurement. *Psychological Bulletin*, 1972, **78**, 64–69.

WINER, B. J. *Statistical principles in experimental design*. (2nd ed.) New York: McGraw-Hill, 1971.

(Received February 5, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 37–43

ROLE OF CENTRAL MONITORING OF EFFERENCE IN SHORT-TERM MEMORY FOR MOVEMENTS

BILL JONES[1]

University of Queensland, St. Lucia, Brisbane, Queensland, Australia

A necessary and sufficient condition for retention of movements over a short time interval may be central monitoring of efferent signals to the muscles (CME) rather than proprioceptive feedback (PFB). It is shown that voluntary movements may be rehearsed over a 15-sec. unfilled retention interval even when location cues are removed, whereas passive movements and constrained movements—movements made to a stop set by *E*, which B. Jones has shown may not be centrally monitored—are subject to decay. Voluntary movements are affected only by an attention-demanding task during the retention interval. Since augmented PFB does not increase accuracy of voluntary movement duplication, CME may be a necessary and sufficient condition for rehearsal in motor short-term memory.

If we are to describe a pattern of movements as skilled, the component movements must at the least be carried out in the correct temporal sequence. Such timing could conceivably depend upon peripheral feedback or may result from central monitoring of efferent commands to the muscles (CME). The present article is concerned with the role of CME compared with proprioceptive feedback from joints and/or muscles (PFB) in short-term retention of single movements (MSTM). Since even the simplest voluntary act requires the regulation of at least 2 antagonistic muscle groups coupled at a joint, timing is a feature of MSTM.

Posner (1967) showed there is little forgetting of visually guided movements over an unfilled retention interval of 20 sec. although an attention-demanding task during the interval did interfere with performance. When *S*s were blindfolded, on the other hand, considerable forgetting occurred after 20 sec. and the interpolated task had no further effect. Posner (1967) argued that central processing capacity is available to the visual channel which functions as a "rehearsal" mechanism and that absence of such capacity makes little difference to the retention of kinesthetic information, even if the interpolated task requires movements as in writing (Posner & Konick, 1966).

However, some movements, including replications of the standard movement during the retention interval (Stelmach & Bassin, 1971), do interfere with MSTM for blind movements. Keele (1968), therefore, suggested that duplication of some movements may be based upon PFB from the muscles, which is not rehearsed, whereas input from the joints has access to central storage and hence may be subject to interference. He did not consider whether CME may be involved, and his PFB argument rests in any case on 2 invalid assumptions.

First, Keele (1968) assumes that only afferents from joint receptors project to the sensory cortex. This is not the case in the monkey and the cat and there is no reason to think that it is the case in man (see Paillard & Brouchon, 1968). The second implied assumption is that the joint receptors signal direction and extent of voluntary movement. However, while we know that joint receptor impulses mediate knowledge of *passive* movements and that the muscle spindles are not involved here (see Rose & Mountcastle, 1959), there is clear evidence that voluntary movements may be accurately duplicated when the joint capsule is anesthetized, thus eliminating joint feedback (Browne, Lee, & Ring, 1954). We can argue, therefore, either that the duplication of voluntary

[1] Requests for reprints should be sent to Bill Jones, who is now at the Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

movement is based upon CME or that PFB from the muscle spindles is tied in with voluntary but not with passive movement (Browne et al., 1954; Paillard & Brouchon, 1968).

Jones (1972) attempted to test the alternative hypothesis by comparing the ability of blindfolded *S*s to produce a voluntary movement based upon one of 3 initial movements. Either (*a*) *S* made a voluntary movement, (*b*) *S* moved to a stop set by *E* ("constrained movement"), or (*c*) *S*'s arm was passively moved to the stop ("passive movement"). There was little difference in voluntary duplication of constrained and passive movements, which were duplicated with significantly less accuracy than were voluntary movements. Jones argued that only in the voluntary condition could *S* predict the end point of the initial movement before he began to move, so that only here was retention mediated by CME. Since both voluntary and constrained conditions required maintenance of muscle tone during the initial movement, there is no reason to think that the behavior of muscle spindles in the 2 movements would have differed; and yet without the opportunity to make a preset movement, reproduction accuracy decreased to the level of passive movement. It may be that both passive movement and constrained movement are based upon joint inflow, or that duplication of constrained movement requires such a complex integration of CME over sections of the movement that the initial movement did not lay down an adequate trace. In any case, there is evidence against the hypothesis of spindle involvement in voluntary movement and in favor of the CME position.

The hypothesis tested in Experiment I is that the initiation of a voluntary movement results in a central record of efferent discharge, or "efferent copy" (von Holst, 1954), which is in itself a motor memory storage system without the requirement of peripheral feedback. Since the efferent copy is central it may be subject to interference, unlike PFB which has no access to central mechanisms. In other words, CME is a necessary condition for rehearsal in MSTM.

Experiment I

Most previous studies of motor retention have had *S*s duplicate a constrained movement and so may have provided little opportunity for retention via CME. The first experiment compared retention of voluntary, constrained, and passive movements over both filled and unfilled retention intervals. The main measure of movement reproduction accuracy was variable error (VE)—mean within-*S*s standard deviation about the algebraic mean. Many other studies of movement duplication (including my own) have had absolute error (AE) as the dependent variable. However, AE confounds VE with algebraic or constant error (CE) (Laabs, 1973; Pepper & Herman, 1970).

Method

Subjects. The *S*s were 24 undergraduates in psychology who took part in the experiment as a course requirement.

Apparatus. A freely moving slide on which *S* placed his index finger was mounted on a straight track. The slide could be moved along the track by either *S* or *E* since a small handle was fixed to the slide on the opposite side to where *S* sat and could be gripped by *E* and used to move the slide with *S*'s arm relaxed. A pointer fixed to the slide moved along a scale so that linear displacements of the slide could be measured. Two freely moving stops were mounted on the track so that position and extent of displacement could be controlled by *E*. A stopwatch was used to time retention intervals.

Procedure. The *S* sat facing *E*, and the different conditions were explained to him. The *S* was then blindfolded. Three movement conditions—voluntary, constrained, and passive—were performed under conditions of 0-sec. delay, 15-sec. rest interval, and 15-sec. task interval. For voluntary movement *S* placed his index finger on the slide, moved it along the track to whatever position he chose, and immediately returned to the initial starting position. The *S*s were instructed to vary the extent of the initial displacement over trials between approximately 10 and 60 cm. Following the retention interval *S* attempted to voluntarily duplicate the initial movement. In the constrained movement conditions, *S* moved the slide to a stop set by *E* and again returned the slide to the initial starting point. During the retention interval *E* removed the stop which indicated terminal position. In the

TABLE 1

Variable, Absolute, and Constant Error (in centimeters) of Movement Duplication in Experiment I

Retention interval	Voluntary			Constrained			Passive		
	VE	AE	CE	VE	AE	CE	VE	AE	CE
0 sec.	1.48	1.81	−.38	3.40	3.15	1.20	2.99	3.39	−2.25
15 sec. rest	1.29	1.65	0.00	4.98	5.20	2.00	4.74	5.90	1.35
15 sec. task	3.93	4.10	3.10	4.53	5.39	2.61	5.27	5.55	−.46

Note. VE = variable error, AE = absolute error, CE = constant error.

passive condition, *S* placed his index finger on the slide and relaxed his arm and *E* then removed the slide to the initial starting point. The *S* always used his preferred arm and initial displacement was always an extensor movement. In all conditions *Ss* were instructed to make both the initial displacement and the attempted duplication as quickly as possible.

Attempted duplication was made either immediately upon returning to the starting point (0-sec. delay) or after 15 sec. The *S* either simply sat with his index finger on the slide for the retention interval until given the instruction "Move" (15-sec. rest), or he was required to count backwards from a 3-digit number spoken by *E* as soon as *S* had returned to the starting point (15-sec. task).

Both movement and interval conditions were within-*Ss* variables counterbalanced in ABCCBA fashion. That is, the order of blocks of movement conditions was voluntary, constrained, passive, passive, constrained, voluntary, and within each block order of retention intervals was 0-sec. delay, rest, task, task, rest, 0-sec. delay, yielding 36 cells each of 6 trials (216 trials overall). To insure approximately equal distributions of initial displacements across conditions, settings in the constrained and passive conditions followed initial settings for each *S* of the settings he had made in the corresponding cell of the first block of voluntary trials. For example, initial displacements made during the second presentation of 6 trials in the 15-sec. rest condition in the first block of voluntary trials were used as standards for duplication in the second presentation of 15-sec. rest in both blocks of constrained and passive movement trials.

Results and Discussion

There were no significant sequence effects and mean VE, AE, and CE for each condition are given in Table 1. A 3 × 3 analysis of variance (ANOVA) of VE with both main effects as within-*Ss* variables showed highly significant effects for movement conditions, $F\ (2, 46) = 157.56$, $p < .000001$, and interval conditions, $F\ (2, 46) = 80.32$, $p < .000001$, and a highly significant interaction term, $F\ (4, 92) = 14.52$, $p < .000001$. Duncan's multiple range test indicated a significant difference ($p < .01$) between the voluntary condition and the other 2 movement conditions with 0-sec. delay. At this interval passive and constrained movement did not differ, thus confirming previous findings (Jones, 1972) in terms of VE. There was no significant difference in either the constrained or the passive condition between rest and task intervals. In the voluntary condition, on the other hand, only the interpolated task produced significant forgetting ($p < .01$), so that the pattern of results for voluntary movement compares with Posner's (1967) findings for visually guided movements.

Figure 1 is a plot of VE against initial displacement. The *S*'s initial displacements were grouped to the nearest 5 mm. and averaged. Inspection of the figure shows no evidence of any tendency for error to vary with extent of movement. Since, however, there was an unequal number of observations for each point on the graph, no statistical check has been attempted.

Following Posner's (1967) criteria it is clear that retention and duplication of voluntary movement may depend upon central processing, since there was no forgetting of voluntary movements over the rest interval while the attention-demanding task interfered with retention.

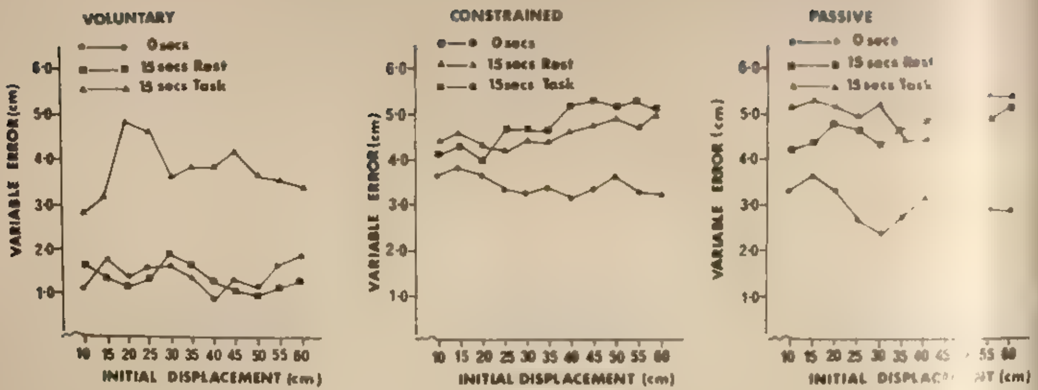


FIGURE 1. Plot of variable error against initial displacement (Experiment 1).

In the constrained and passive conditions, on the other hand, forgetting occurred over time and there was no interference effect. Since only the voluntary condition allowed *S* to preset the initial movement, the present experiment provides evidence that CME rather than joint PFB is a necessary condition for motor rehearsal. Keele and Ells (in press) have recently shown that passive movements may show little forgetting in the absence of distraction, indicating that efferent consistency may not be required for motor rehearsal. However, they carried out passive movement at a constant and steady rate, which may have been highly susceptible to a counting strategy (Laabs, 1973, has shown that counting is a very efficient way of reducing VE in movement duplication), unlike the fast rate of movement used in the present experiment. Consequently, performance in the Keele and Ells study may not have depended upon joint PFB.

Against the present position it might be argued that when *S*s are asked to make a voluntary movement they decide on a stopping position and later remember that position rather than efference from the movement. The second experiment was designed to test such a possibility.

EXPERIMENT II

Method

Subjects. The *S*s were 15 students who had not taken part in Experiment I.

Apparatus and procedure. The apparatus was that used in Experiment I. The *S*s carried out voluntary movement duplication under 2 conditions. The first (constant start) followed exactly the voluntary movement procedure in the first experiment. The second (variable start) was the same as the first except that the starting position of the standard movement and of the attempted duplication were randomly varied on each trial over 6 randomly chosen positions. The second, third, fourth, fifth, and sixth positions were, respectively, extensions of 9, 13, 25, 32, and 45 cm. from the first position, the midline of *S*'s body. Consequently, *S* had to duplicate the extent of the initial displacement, and without the possibility of location cues.

Both movement conditions and interval conditions were within-*S*s variables counterbalanced as in Experiment I. Each *S* received 240 trials overall in blocks of 10 each.

Results and Discussion

Table 2 gives mean VE and mean CE for the various conditions. Analysis of variance of VEs showed that only the intervals factor was significant, $F\ (2, 28) = 362.34$, $p < .000001$. Significance was not obtained for either the movements main effect, $F\ (1, 14) = .01$, or the Movements $\times$ Intervals interaction, $F\ (2, 28) = 1.40$. Duncan's multiple range test showed that differences between the 0-sec. and 15-sec. tasks were significant ($p < .01$) for both variable and constant start conditions, although for neither was there a significant difference between 0-sec. and 15-sec. rest. Experiment II consequently provides evidence that MSTM for blind voluntary movements follows the pattern of retention of visually guided movements (Posner, 1967) even when location cues are eliminated.

TABLE 2

Variable Error (in centimeters) and Constant Error (in centimeters) of Movement Duplication in Experiment II

Retention interval	Constant start VE	Constant start CE	Variable start VE	Variable start CE
0 sec.	1.34	−.51	1.43	−.44
15-sec. rest	1.36	−.47	1.55	−.60
15-sec. task	4.46	2.79	4.14	3.73

Note. VE = variable error, CE = constant error.

Experiment III

In general the results of Experiments I and II are good evidence that CME is a necessary condition for rehearsal in MSTM processes. So far the evidence does not show whether CME is both necessary and sufficient. In bilaterally deafferented monkeys CME certainly seems to be sufficient for motor learning (Taub & Berman, 1968), but in intact human *Ss* it may be that PFB "stamps in" the efferent copy. The present experiment was designed to test the hypothesis that PFB facilitates voluntary movement duplication since, if this is the case, CME would be a necessary but not a necessary and sufficient condition for retention of a voluntary movement over time. By having *S* move the slide against varying degrees of tension, PFB was manipulated. Such a recourse has frequently been held to increase PFB (e.g., Bahrick, 1957) and, since Pepper and Herman (1970) have reported that MSTM for force is analogous to MSTM for extent and position, it is reasonable to expect that increasing force information could lower VE of voluntary movement duplication, given that PFB interacts with CME to stabilize memory for extent of movement.

Method

Subjects. The *Ss* were 15 undergraduates who had not taken part in either Experiment I or Experiment II.

Apparatus. The apparatus was the slide and track system used previously, with the addition that a weight could be attached to the slide by a cord and pulley 10 cm. in diameter mounted at one end of the track.

Procedure. The general procedure followed that already described for the variable start condition in Experiment II, with differential weight on the slide of 0, .5, 1.0, and 1.5 kg. Weight was a within-*Ss* variable counterbalanced across trials as were retention intervals. Each *S* received 240 trials in all.

Results and Discussion

Table 3 gives mean VE and mean CE for each condition. A 4 × 3 ANOVA of VEs with weights and intervals as within-*Ss* factors showed, as for Experiment II, that only the main effect of intervals was significant, $F\ (2, 28) = 329.27$, $p < .000001$. Neither the effect of differential weighting, $F\ (3, 42) = 1.77$, nor the Weights × Intervals interaction, $F\ (6, 84) = .42$, reached significance.

Breakdown of the analysis of variance by Duncan's test did not indicate significant differences between movement conditions at any of the interval levels. Thus the present findings for voluntary movements parallel those of Stelmach (1968) for con-

TABLE 3

Variable Error (in centimeters) and Constant Error (in centimeters) of Movement Duplication in Experiment III

Retention interval	0 kg. VE	0 kg. CE	.5 kg. VE	.5 kg. CE	1 kg. VE	1 kg. CE	1.5 kg. VE	1.5 kg. CE
0 sec	1.57	0.00	1.46	−.74	1.35	−.55	1.37	−.50
15-sec. rest	1.54	.17	1.49	−.36	1.45	−.68	1.37	−.37
15-sec. task	4.21	2.68	4.33	3.47	3.99	2.74	4.17	2.58

Note. VE = variable error, CE = constant error.

strained movements in that augmenting PFB without also giving visual or verbal knowledge of results has little effect on accuracy of movement duplication. Consequently, there is evidence for the hypothesis that CME is both a necessary and sufficient condition for rehearsal in MSTM. Limiting factors on the present conclusion are that force and extent of movement may be subserved by different neural mechanisms despite Pepper and Herman's (1970) findings, and also that VE in the 0-kg. condition may represent the limits of accuracy at present levels of practice so that augmenting PFB could not facilitate motor retention. Further experimentation could study the effects of various tension levels of more long-term practice on blindfolded voluntary movement duplication, since it may be that augmented PFB results in rather faster lowering of VE over a longer time span.

General Discussion

The present results are good evidence that CME is a necessary condition for rehearsal in MSTM, and they also provide a suggestion that CME is necessary and sufficient. Most if not all previous studies of MSTM have failed to consider the CME hypothesis since they have compared the duplication of constrained movements under various conditions. Peripheral feedback theories of MSTM have therefore been given a generality on the basis of insufficient evidence. This is not to say that memory storage capacity for proprioceptive inputs does not exist. It is simply that MSTM for proprioceptive inputs need not involve central mechanisms since storage may be via "preperceptual" kinesthetic afterimages comparable to the sensory afterimages found in vision and hearing, although probably with a relatively longer decay function. It is largely meaningless to speak of interference here; any new inputs to the preperceptual store would erase existing afterimages rather than interacting in classical interference fashion. It is conceivable that voluntary movement also results in an afterimage which may or may not be read in. The absence of any proprioceptive facilitation effect on duplication of voluntary movement would suggest that, if it occurs, preperceptual storage of data from proprioceptors resulting from voluntary movement has little functional utility. The adaptation theory developed by Pepper and Herman (1970) and extended by Laabs (1973) seems to give a good fit to the data on short-term retention of constrained movements. However, such a model may have little extension to the wider field of skilled performance. Ordinarily skill means a pattern of correctly timed *voluntary* movements, and the present data show that voluntary and constrained movements differ in important ways. In skilled performance, generally, the initial learning stages may depend upon visual (or other exteroceptive) feedback matched against the motor memory or efference copy to specify error of movement. With increasing practice *S*s may come to depend solely on CME, contrary to the widely held theory (e.g., Bahrick, 1957) that PFB control replaces visual timing of movements.

The neural basis of CME possibly depends upon cerebello-cerebral relationships. It is known that coordinated movements are disrupted by cerebellar damage even though the motor cortex remains intact. Ruch (1951) speculates that preprogramming of movements would be possible if the motor cortex initially discharges into a circular pathway involving the cerebellum, so that impulses would be stored to be discharged after a fixed delay. Such a "reverberatory circuit" could conceivably mediate memory for voluntary movements over the 15-sec. rest interval in the present experiment.

References

Bahrick, H. An analysis of the stimulus variables influencing the proprioceptive control of movements. *Psychological Review*, 1957, 64, 324–328.

Browne, K., Lee, J., & Ring, P. A. The sensation of passive movement at the metatarso-phalangeal joint of the great toe in man. *Journal of Physiology*, 1954, 126, 448–458.

Jones, B. Outflow and inflow in movement duplication. *Perception & Psychophysics*, 1972, 12, 95–99.

Keele, S. W. Movement control in skilled motor performance. *Psychological Bulletin*, 1968, 70, 387–403.

Keele, S. W., & Ells, J. G. Memory characteristics of kinesthetic information. *Journal of Motor Behavior*, in press.

Laabs, G. J. Retention characteristics of different reproduction cues in motor short-term memory. *Journal of Experimental Psychology*, 1973, 100, 168–177.

Paillard, J., & Brouchon, M. Active and passive movement in the calibration of the position sense. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior*. Homewood, Ill.: Dorsey Press, 1968.

Pepper, R. L., & Herman, L. M. Decay and interference effects in short-term retention of a discrete motor act. *Journal of Experimental Psychology Monograph*, 1970, **83** (2, Pt. 2).

Posner, M. I. Characteristics of visual and kinesthetic memory codes. *Journal of Experimental Psychology*, 1967, **75**, 103–107.

Posner, M. I., & Konick, A. F. Short-term retention of visual and kinesthetic information. *Organizational Behavior and Human Performance*, 1966, **1**, 71–88.

Rose, J. E., & Mountcastle, U. B. Touch and kinesthesis. In J. Field (Ed.), *Handbook of physiology*. Section 1. *Neurophysiology*. Vol. 1. Washington, D.C.: American Physiological Society, 1959.

Ruch, T. C. Motor systems. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Stelmach, G. E. The accuracy of reproducing target positions under various tensions. *Psychonomic Science*, 1968, **13**, 287–288.

Stelmach, G. E., & Bassin, S. L. The role of overt motor rehearsal in kinesthetic recall. *Acta Psychologica*, 1971, **35**, 56–63.

Taub, E., & Berman, A. J. Movement and learning in the absence of sensory feedback. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior*. Homewood, Ill.: Dorsey Press, 1968.

von Holst, E. Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behaviour*, 1954, **2**, 89–94.

(Received March 1, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 44–49

INFORMATION AMPLIFIED:

MEMORY FOR COUNTERFACTUAL CONDITIONALS[1]

SAMUEL FILLENBAUM[2]

University of North Carolina at Chapel Hill

People remember not only information explicitly specified but also information that can be inferred from that provided directly. It was found, in a recognition memory task that compared to control items *S*s "recognized" the negated antecedent and consequent propositions of previously encountered counterfactual conditionals significantly more often, the latter effect being distinctly stronger. A similar result was obtained for causals related to previously encountered counterfactual conditionals and counterfactual conditionals related to previously encountered causals, the latter being the stronger effect. These results were evaluated in the context of the observation that a counterfactual conditional presupposes (*a*) the negation of its antecedent proposition, strongly suggests (*b*) the negation of its consequent proposition, and (*c*) is intimately related to a causal in which *a* and *b* are conjoined, with *a* taken as the grounds for *b*.

Characteristically, people remember neither the particular surface syntactic form of the sentences they have encountered nor even their deep structure, but rather remember semantic information or propositional content that includes not only information explicitly specified but also information that can be inferred from that provided directly (for a survey of relevant evidence see Fillenbaum, 1973). Thus, Tieman (1971), for example, has shown that if a sentence pair such as *John is tall John is shorter than Bill* is presented, *S*s may later often falsely recognize *Bill is tall*, which can be inferred from the actually presented sentences. Bransford, Barclay, and Franks (1972) provide related evidence in studies concerned with both memory for inferences that may be drawn from sets of semantically related sentences and recognition memory for information inferable from single sentences, given the proper understanding of lexical items such as [illegible] proposition.

These examples represent relatively simple cases of information compounding or the drawing of simple 1-step inferences. There is very little knowledge of what may happen when more subtle types of inferences are involved, even for the case of single sentences, as regards the recall of what is either presupposed or possibly entailed by a sentence, as compared to recall for what is explicitly asserted in it. The research presented herein is addressed to these questions. Recently, linguists have become very interested in presuppositional analysis (see, e.g., Fillmore & Langendoen, 1971). Psychologists have begun to consider how to differentiate between the knowledge involved in what is explicitly asserted and what is presupposed in an expression and to examine the role of knowledge involved in such presuppositions (see, e.g., Miller, 1969; Osgood, 1971). If in processing information one is responsive to not only what is explicitly and directly provided but also to what can or must be assumed if the directly provided information is to be coherent, then a concern with the functions of the presuppositional knowledge becomes inescapable. Particularly intriguing questions arise with regard to possibilities of information gain or amplification in memory, as opposed to the more usual

[1] This research was supported in part by U.S. Public Health Service Grant [illegible] from the National Institute of Mental Health. The author is indebted to K. H. [illegible] for [illegible] gathering and analyzing the data.

[2] Requests for reprints should be sent to Samuel Fillenbaum, L. L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill, North Carolina 27514.

emphasis on abstraction and information loss.

The conceptual status and proper analysis of counterfactual conditionals have long concerned and perplexed philosophers; nevertheless counterfactual conditionals may provide us with test material that is particularly useful and relevant for the study of memory for information implicit in sentences. A counterfactual conditional presupposes the negation of the antecedent proposition and, while there is some dispute as to whether or not it also entails negation of the consequent proposition, it would seem that it (at the least) strongly suggests such negation and "invites" this as an inference (see Karttunen, 1971). Thus, given *If he had caught the plane he would have arrived on time*, it is presupposed that *He did not catch the plane*, and in most contexts it is at least strongly suggested that *He did not arrive on time*. A counterfactual conditional may also be seen as related to a causal in which the negation of the antecedent is given as the reason for the negation of the consequent, in our example *Because he did not catch the plane he did not arrive on time* or, more stylistically common, *He did not arrive on time because he did not catch the plane*.

Given prior presentation of a counterfactual conditional, one may ask if *S* will later falsely recognize (*a*) the negation of the antecedent proposition, which is presupposed or (*b*) the negation of the consequent proposition, which is strongly suggested if not entailed; and whether false recognition is more frequent for *a* or *b*. Further, given a counterfactual conditional, one may ask if *S* will later falsely recognize its causal relative, which consists of the causal conjunction of *a* and *b*, with *a* given as the reason for *b*. This relation between counterfactual conditional and causal may be viewed as going in the other direction as well (i.e., from causal to counterfactual), since there would appear to be few, if any, occasions appropriate for the causal (*He did not arrive on time because he did not catch the plane*) which would not also assume the counterfactual (*If he had caught the plane he would have arrived on time*). Naturally, we shall be interested in comparing the frequency of recognition errors in the 2 directions. Two experiments in recognition memory are presented herein, the first of which is concerned with memory for the simple offspring of counterfactual conditionals and the second with memory for related complex sentences where the original sentences are either counterfactuals or causals and the recognition sentences include their causal and counterfactual relatives.

METHOD

The basic format of the 2 experiments was the [illegible] were told that they would [illegible] were requested to listen carefully, since they would later be [illegible] After the [illegible] series was finished, they were told that they would hear another series of sentences and would then have to decide whether each sentence was *old* (one they had actually heard before) or *new* (one that had not appeared in the first series). The sentences were presented at a 5-sec. rate both in the initial series and in the later recognition series. The *S*s were run in groups ranging in size from 4 or 5 to 20 or so. There were 110 *S*s in Experiment I and 92 *S*s in Experiment II, all undergraduates fulfilling course requirements.

Design.—Experiment I. Both the original and the test series consisted of a mixture of counterfactual conditionals and simple propositions. Whether [illegible] the original or in the test series, all counterfact[illegible] were positive in both of their clauses with the *if* clause always coming first, as in *If he had caught the plane he would have arrived on time*, and all [illegible] *He did not catch the plane*. Some physical or social [illegible] counterfactuals, as in *If he had eaten the fish he* [illegible] *the boss he would have got a raise*. Matters were arranged so that there were (*a*) conditionals appearing both in the original and the test [illegible] (6 items), (*b*) simple propositions appea[illegible] the original and the test series (8 items), [illegible]ditionals appearing only in the test series (12 [illegible]) (*d*) simple propositions appearing only in the test series (8 items), (*e*) conditionals appearing in t[illegible] original series whose negated antecedent appeare[illegible] [illegible] sequent appeared in the test series (8 items), and [illegible] pearing both at the start and end of the [illegible] series and at the start of the test series, as well as scattered through the original series, so as to make the proportions of conditionals and simple

propositions comparable in the original and test series. Each item appeared twice in the original series; after every item had occurred once, the items appeared again in a scrambled order.

Design—Experiment II. The original series consisted of a mixture of counterfactual conditionals, simple conditionals, and causals involving negation in both of their propositions. Simple conditionals were included as a control so as to be able to determine whether any systematic recognition effects obtained were simply topic effects, i.e., were due to the prior occurrence of the same or similar content. The same type of sentence content used in Experiment I appeared in this experiment. In the causals, the *because* clause always came second, as in *He didn't pass the test because he didn't study.* The simple conditionals involved the same kinds of connections between their clauses as did the counterfactuals, e.g., *If he is treated immediately he will survive* or *If the watches the show he will be bored.* Matters were arranged so that there were (*a*) counterfactual conditionals, simple conditionals, and causals appearing both in the original and test series (6 items of each kind), (*b*) counterfactual conditionals, simple conditionals, and causals appearing only in the test series (6 items of each kind), (*c*) counterfactual conditionals appearing in the original series whose related causals appeared in the test series (6 items), (*d*) causals appearing in the original series whose related counterfactual conditionals appeared in the test series (6 items), (*e*) simple conditionals appearing in the original series whose corresponding causals appeared in the test series (6 items), and (*f*) some filler items. Again, each item appeared twice in the original series, and after every item had occurred once the items appeared again in a scrambled order.

For each *S*, one may determine the number or proportion of *old* responses for each item class and then make appropriate comparisons across item [illegible]; in addition, one may determine the mean [illegible] of *old* responses taken over *Ss* for each [illegible] class. These figures will constitute our [illegible] data

RESULTS

Consider first the results for Experiment I. *Old* counterfactuals (.88) and [illegible] were recognized as *old* significantly more frequently than [illegible] and *new* simple propositions (.14)[6]; *old* counterfactuals [illegible] simple propositions (.53); and there was no difference in recognition between *new* counterfactuals (.12) and *new* simple propositions (.14). Except for the finding that *old* counterfactuals are recognized better than *old* simple propositions, these results are of no particular interest except to show that the situation is one in which *S*s can distinguish between *old* and *new* items, and thus one where the results for offspring items will be relevant. It is indeed the case that offspring sentences, whether negated antecedent (.25) or negated consequent (.44), were judged *old* significantly less frequently than *old* simple propositions (.53), but more interestingly, it is also the case that these offspring sentences were judged to be *old* sentences significantly more often than the control *new* simple propositions (.14). Finally, if we compare offspring sentences where the consequent is negated with those where the antecedent is negated, we find that the former (.44) are judged *old* significantly more often than the latter (.25).

In Experiment II, *old* counterfactual conditionals (.79), *old* simple conditionals (.84), and *old* causals (.73) are recognized as *old* more than *new* counterfactual conditionals (.05), *new* simple conditionals (.04), and *new* causals (.08), with each of these differences highly significant. While the differences between *old* propositions are significant, with simple conditionals recognized significantly more often than counterfactual conditionals or causals and counterfactual conditionals recognized more often than causals ($p < .05$, sign test), these differences are quite small compared to those between *old* and *new* propositions. The differences between various sorts of *new* propositions are slight and nonsignificant. Related sentences are judged *old* significantly less often than sentences that actually had appeared before, but more interestingly, related sentences are judged *old* significantly more often than their *new* controls. Further, looking in a little more detail, we find that counterfactual conditionals related to previously presented causals are "recognized" significantly more often (.47) than causals related to previously presented

[6] Unless explicitly noted otherwise, all differences described as significant represent $p < .01$ on a [illegible]

counterfactuals (.34), and that the latter are "recognized" significantly more often than causals corresponding to previously presented simple conditionals (.25). As noted above, *old* counterfactual conditionals are recognized more often than *old* causals (.79 vs. .73), so perhaps the difference between counterfactuals related to previously presented causals (.47) and causals related to previously presented counterfactuals (.34) represents only a bias toward recognition of counterfactuals. Thus, one needs to correct recognition rates for related sentences by corresponding rates for *old* sentences in order to determine whether the difference between .47 and .34 is reliably greater than that between .79 and .73. Comparing for each *S* the difference in recognition rate between *old* counterfactuals and *old* causals with that between related counterfactuals and related causals, we find that the latter difference is greater for a significant majority of *Ss* ($p < .05$). Thus, even with this correction taken into account, counterfactual conditionals related to previously presented causals are recognized significantly more often than causals related to previously presented counterfactuals.

Discussion

The basic trends of the results are quite clear in each experiment and consistent over the 2 experiments. While offspring or related sentences are recognized less frequently than *old* sentences, they are recognized more frequently than *new* control sentences. Listeners appear to operate on the information directly provided to them, amplifying and elaborating that information. In considering the results more closely, one important caution needs to be voiced with regard to their interpretation. There is no statistical warrant for generalizing these results beyond the specific samples of linguistic material actually employed. There are obviously very severe problems in defining a population of counterfactual conditionals and in sampling at random from such a population; for a discussion of the "language as fixed effect fallacy" and some suggestions for corrective measures, see Clark (in press).

In might perhaps be argued that the positive findings of Experiment I merely represent some sort of a topic effect, with negated antecedent and consequent propositions falsely recognized because *S* remembers having encountered some similar content in the original series (where these contents were actually embedded in various counterfactual conditionals). This argument would not appear to be too convincing given the findings of Experiment II, where there was significantly more false recognitions of causals related to prior counterfactual conditionals than of causals corresponding to prior simple conditionals. If all that we have is a topic effect, too, there should have been no difference in false recognition of causals, whether the antecedent conditionals were simple or counterfactual. Nor would the topic hypothesis account in any obvious way for the fact that in Experiment I, false recognition of negative consequents was significantly more frequent than that of negative antecedents. Why should this be the case if all that is involved is the recognition of propositions that had occurred earlier in some form? Finally, note that recognition involved the negation of the explicit prior contents, and we know that gist memory is generally good, so that on recall *S* can discriminate rather well between an expression and its negation. Nevertheless, a built-in control against the topic hypothesis may be desirable in future work, and one suggests itself readily. In the recognition series, in addition to sentences which negate the antecedent and consequent propositions of earlier counterfactual conditionals, which we might call "legitimate" offspring, one might also include sentences which assert the antecedent and consequent propositions. These might be called "illegitimate" offspring. On the basis of a topic hypothesis, no difference should be expected in false recognition of the legitimate and illegitimate offspring, or perhaps we might expect even greater recognition of the latter than the former (since the illegitimate offspring do not involve negation of previously encountered materials). In terms of the position argued here, however, one would expect more false recognition of legitimate than of illegitimate offspring sentences.

With regard to the first experiment, we can offer no plausible reason as to why old counterfactual conditionals should have been recognized substantially more often than old simple propositions (.88 vs. .53) except for noting that there were substantially fewer of the former (18) than of the latter (32) in the recognition series, and this may have

resulted in less confusion for the former. A problem more directly related to the concern of the present research is posed by the finding that sentences which negate the consequent of a counterfactual conditional are significantly more frequently recognized than those which negate the antecedent of the counterfactual conditional (.44 vs. .25).[4] Recall that the negation of the antecedent is strictly presupposed by the counterfactual and that one cannot coherently conceive of a counterfactual without assuming that its antecedent proposition is negated, while it has been argued that the negation of the consequent is strongly suggested in most circumstances but does not appear to be required. Thus, a sentence such as *If he had caught the plane, which he did, he would have arrived on time* seems inconsistent, while there is in principle nothing contradictory in saying *If he had caught the plane he would have arrived on time, which he did anyway because he took a fast train.* Hence it would appear that, if anything, the negated antecedent is more strongly implied by the counterfactual conditional than is the negated consequent, and yet our results indicate substantially more false recognition of the latter than of the former. In order to make sense of this finding, consider the circumstances under which one might be most likely to employ a counterfactual conditional. For example, one might most commonly say *If he had caught the plane he would have arrived on time* as an explanation for the fact that he was late, a fact that may be simply expressed by negating the consequent proposition, i.e., saying *He did not arrive on time.* If the fact that is being explained is somehow central or focal to the discourse, then it may be reasonable that a proposition that embodies this fact, i.e., the denied consequent, should be more likely to be inferred and falsely remembered than one that only purports to give a reason for the fact, viz., the subordinate denied antecedent.[5] All this, while perhaps sensible enough, is of course somewhat ad hoc, since one could argue that the subordinate causal, just because it gives an explanation, is the more central proposition. What is needed is some manipulation that might systematically change the focus between the 2 propositions, and a determination of whether such a shift in focus has any systematic effect on recognition errors in a task of the sort used in the present study.

With regard to the second experiment, it is again hard to know why there should be differences in recognition of the various kinds of *old* sentences, but here the differences are certainly not of any great magnitude with figures of .79, .84, and .73 for counterfactual conditionals, simple conditionals, and causals, respectively. The finding that causals corresponding to simple conditionals are recognized more often than *new* causals suggests that in some measure there may indeed obtain a sort of topic effect, such that a sentence involving propositional material encountered earlier may be falsely recognized. However, the finding that causals related to previously encountered counterfactual conditionals are significantly more frequently falsely recognized than causals corresponding to previously encountered simple conditionals indicates that such a topic effect cannot totally explain present results. Perhaps the most interesting result is the finding that both causals related to counterfactual conditionals and counterfactual conditionals related to causals are recognized significantly more often than their controls, with the latter effect the stronger one. But why should the counterfactual relative of a causal be more frequently recognized than the causal relative of a counterfactual? Consider what is assumed or what can be assumed given, respectively, a causal and its related counterfactual conditional, e.g., the sentences (*a*) *He did not arrive on time because he did not catch the plane* and (*b*) *If he had caught the plane he would have arrived on time.* One cannot very easily assert *a* without assuming or taking for granted the correctness or truth of *b*, but it is not clear that the converse necessarily holds. For, as was pointed out earlier, *b* could be true and yet there is no absolute requirement that the person in question must have been late, for (*c*) *If he had*

[4] Actually, recognition memory for offspring sentences was examined both for the case where just 1 offspring of a particular counterfactual conditional was presented in the recognition series (with a negated antecedent provided on half of the occasions and a negated consequent on the other half) and for the case where both the negated antecedent and the negated consequent offspring of a counterfactual were presented. Since it made essentially no difference whether one or both offspring were tested, only the pooled results are presented.

[5] Another possibility, pointed out by a reviewer, is simply that semantic material from the main clause is better remembered than material from the adverbial *if* clause.

caught the plane he would have arrived on time, which he did anyway because he caught a fast train, is certainly a possible elaboration. The point is simply that a counterfactual conditional does not appear to entail the denial of the consequent proposition and thus, insofar as this is the case, does not require its related causal. But this analysis too may appear ad hoc. At least it would be supportive of these suggestions if, given causals and their counterfactual relatives, *S*s were more willing to accept the latter as inferences on the basis of the former than the converse.

Even though our interpretation of some of the details of the results must be quite tentative with regard to within-conditions differences, viz. (*a*) the difference between offspring sentences which negate the antecedent and those which negate the consequent and (*b*) the difference between causals related to counterfactual conditionals and counterfactual conditionals related to causals, it seems clear that there are substantial and consistent between-conditions differences, such that offspring sentences and related sentences are consistently recognized more often than their *new* controls. This occurs with regard to both simple offspring propositions (Experiment I) and complex related propositions involving the causal conjunction of the simple offspring (Experiment II).

Wittingly or unwittingly, *S*s seem to go beyond the givens, using information explicitly provided as a clue to what is implicit or can be assumed given that information. If this is the case, then it is incumbent on us to investigate the processes of inference or natural logic involved in such elaboration of information.

REFERENCES

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 1972, **3**, 193–209.

Clark, H. H. The language-as-fixed-effect-fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, in press.

Fillenbaum, S. *Syntactic factors in memory?* The Hague, The Netherlands: Mouton, 1973.

Fillmore, C. J., & Langendoen, T. (Eds.) *Studies in linguistic semantics.* New York: Holt, Rinehart & Winston, 1971.

Karttunen, L. Counterfactual conditionals. *Linguistic Inquiry*, 1971, **2**, 566–569.

Miller, G. A. A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 1969, **6**, 169–191.

Osgood, C. E. Where do sentences come from? In D. D. Steinberg & L. A. Jacobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology.* Cambridge: Cambridge University Press, 1971.

Tieman, D. G. Recognition memory for comparative sentences. Unpublished doctoral dissertation, Stanford University, 1971.

(Received March 3, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 50–55

MULTIPLE-EXPOSURE EFFECTS ON RESOLUTIONS OF FOUR BASIC CONFLICT TYPES[1]

BARRY D. SMITH[2]
University of Maryland

LEONARD GEHL
New Jersey College of Medicin

The present study examined the 4 Lewinian conflicts in the context of a negative-start paradigm. The *S*s were given 3 trials of each of the 4 conflicts, interspersed among 60 nonconflict trials. Double approach–avoidance (Type IV) conflict produced less adequate resolutions than avoidance–avoidance (Type III), but the 2 conflicts were similar in response modes employed, slow speed of resolution, and high error rate. In addition, considerably less blocking was found to Type III and IV conflicts than in work performed by Hovland and Sears in 1938. Repeated exposure to conflicts produced increases in speed. Resolution lability took the form of increased use of double and blocking responses and decreased use of single responses across trials and was affected by conflict type and response adequacy. Approach–avoidance conflict showed the greatest overall lability, while interclass shifts were greatest for double approach–avoidance.

Two recent studies of human cognitive–motor conflict (Epstein & Smith, 1967; Smith & Epstein, 1967) analyzed responses to approach–approach (Type I), approach–avoidance (Type II), and avoidance–avoidance (Type III) conflict. These studies modified an experimental paradigm used earlier by other investigators (Hovland & Sears, 1938; Sears & Hovland, 1941) to include a negative-start condition in which *S*s were instructionally forced to leave the starting point in the conflict field immediately at the onset of any conflict or nonconflict trial. Results of the Epstein–Smith and Hovland–Sears studies were quite discrepant on some points. In particular, Hovland and Sears found that blocking (remaining at the starting point after trial onset) was a common response to Type III conflict, while Epstein and Smith, using the negative-start paradigm, found blocking to be quite infrequent.

Although the Epstein–Smith (Epstein & Smith, 1967; Smith & Epstein, 1967) studies did introduce the negative-start paradigm, they limited the examination of conflict responses to the first 3 conflict types, eliminating the double approach–avoidance (Type IV) conflict examined by Hovland and Sears (1938), and to a single presentation of each conflict, with the 3 conflict trials interspersed among nonconflict trials. Hovland and Sears did study the resolution of Type IV conflict, but their results were undoubtedly affected by their failure to use a negative starting point. In addition, comparisons between Type IV and the other 3 types in the Hovland and Sears study are difficult, since they used only 20 preliminary trials with each of the first 3 conflict types and 80 with Type IV. The present study was carried out, in part, to evaluate Type IV conflict effects in comparison with the other conflict types.

Hovland and Sears (1938) also studied the effects of repeated exposure to conflict, but only with regard to Type II conflict. A major purpose of the present study was thus to examine, in the context of the negative-start paradigm, response lability (the tendency to change or shift from one response mode to another with repeated exposure to conflict) for all conflict types.

METHOD

Subjects

A total of 52 Introductory Psychology students (24 male, 28 female) were employed as *S*s. Additional *S*s (7) were run or partially run but had to be dropped due to failure to understand basic

[1] The authors are indebted to Frank Rosillo for his help in running *S*s in the study. Computer time was provided by the Maryland University Computer Science Center.

[2] Requests for reprints should be sent to Barry D. Smith, Department of Psychology, University of Maryland, College Park, Maryland 20742.

instructions (2), overt refusal to follow instructions (1), or equipment failure (4).

pparatus

ne conflict apparatus consisted of a board on ich were mounted 4 lights (one red and one ite light at each of 2 goals), a control panel that rmitted *E* to switch on any light or any 2 or re lights simultaneously, and a Standard electric ier wired to start simultaneously with light onset d to be stopped with the termination of *S*'s ponse. The response field was a simple maze, nsisting of a diamond-shaped section followed by a V-shaped section, with paths terminating at 2 goals. The effect of the maze was such that the first section directed *S*'s vision toward one goal, the second toward the opposite goal, and the third section again toward the first goal. Crossways between the sides of the maze permitted *S* to detour from one side to the other during his response. The actual maze was one presented in an earlier study by Epstein and Smith (1967). The only major difference in the present study was that the outside pathways of the maze were narrowed to 2 mm. to increase error frequency and thus make error a meaningful measure.

Procedure

Each *S* was instructed to approach white lights, which he was told represented a million dollars, and to avoid red lights, representing an atomic bomb, by moving a pencil along the paths of the maze. The starting point was made negative by indicating that it would represent a bomb detonation point whenever any light indicating trial onset came on.

The experimental session consisted of a total of 72 trials, including 3 trials for each conflict type, interspersed among 60 nonconflict (simple approach or avoidance) trials. More specifically, after 14 nonconflict trials, 3 different random sequences of 5 trials each were interspersed among nonconflict trials. Each sequence contained all 4 conflicts and a randomly selected control (nonconflict) trial. All *S*s received the same resulting random schedule of conflict and nonconflict presentations. Type I conflict involved the presentation of the 2 white lights simultaneously, Type II consisted of the red and white lights at the same goal appearing simultaneously, Type III was 2 red lights, and Type IV was all 4 lights. A new copy of the response maze was used for each successive trial.

Results

The results of the study were analyzed in terms of response speed, errors made in responding, mode and adequacy of conflict resolution, and response lability as a function of conflict repetition. All analyses involved all 52 *S*s except where specifically noted.

Response Speed and Error

Effects of conflict type and practice were examined through analyses of the rate at which *S*s traversed the maze in response to a given conflict presentation and the relative frequency of errors made. Analyses of rate involved a speed score, calculated as the ratio of total response time on a given trial to distance traversed. The speed score was calculated for each *S* on each of the 3 presentations of each conflict type and on each of the 3 corresponding random control (nonconflict) trials.

A repeated-measures analysis of variance applied to speed scores on the conflict and control trials, with conflicts and trials as sources of main effect variance, yielded a significant trials main effect, F (2, 56) $= 13.71$, $p < .001$. Speed increased across trials, with mean values of 4.66, 5.18, and 5.43 mm/sec for combined Conflict and Control Trials 1, 2, and 3, respectively. Conflict type was also significant, F (4, 112) $= 30.46$, $p < .001$ (Figure 1). Speed decreased with increasing conflict difficulty except for a slight increase on Type IV. A Scheffé (1959) test indicated that the difference between Type III and Type IV conflicts was not significant, F (4, 112) $= 1.85$, $p > .05$. The difference between Type I conflict and control trials was also clearly nonsignificant.

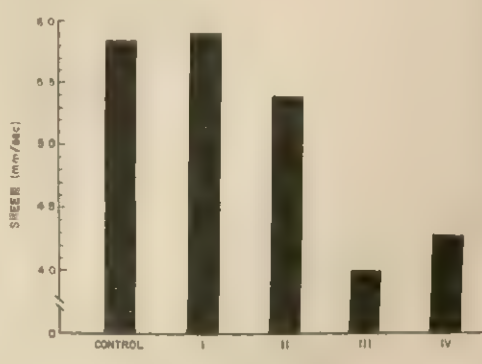


FIGURE 1. Speed of response as a function of conflict type.

A test was planned of the expectation, drawn from the work of Worell (1967), that response speed decreases with repeated exposure to Type IV conflict. The necessary planned comparison was carried out as a Scheffé (1959) test. Means for Trials 1 and 3 on Type IV conflict were 3.83 and 4.71, respectively, representing a significant *increase* in speed rather than the hypothesized decrease, F (8, 224) = 13.51, $p < .10$. A probability of .10 was considered significant due to the highly conservative nature of the Scheffé test.

For the error analyses, an error was defined as any movement of S's response line that crossed the boundary defining a pathway of the response maze. The measure utilized was error rate (errors/speed). The significant effect of conflict type, F (4, 112) = 5.52, $p < .001$, resulted from an increase in error rate across the 4 major conflict types (Figure 2).

Response Mode and Adequacy

Response mode and response adequacy were scored, using the criteria of earlier studies (Epstein & Smith, 1967; Smith & Epstein, 1967). The modes scored were: (*a*) single response, in which the goal is reached directly, without detour; (*b*) double response, in which both goals are reached either by going from one goal to the other or by returning to the starting point and proceeding to the second goal; (*c*) compromise, in which the response terminates at a point intermediate between the starting point and the 2 goals; (*d*) detour, in

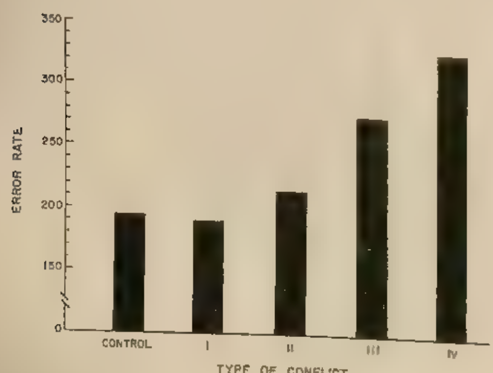


FIGURE 2. Error rate as a function of conflict type.

which the response includes proceeding through one or more crossways of the maze; (*e*) blocking, which is simply a failure to leave the starting point following stimulus onset; and (*f*) disorganized responses, in which no attempt is made to remain within the confines of the maze. Adequacy was scored in relation to specific conflict types. Detour, blocking, and disorganized responses were always scored as inadequate. A compromise response was adequate only for Type III or Type IV conflict, and the double response was adequate only for Type I conflict. Single responses were adequate in Type I conflict, in Type II conflict (to the nonconflicted goal), and on control trials (to the correct goal).

Conflict and corresponding control trials were scored for mode and adequacy, and appropriate percentages were determined. Results of the present study with regard to the first 3 conflict types were in close agreement with the earlier work of Epstein and Smith (1967). Dominant response modes for each conflict type were similar to those in the earlier work, and adequacy decreased with increasing conflict difficulty from Type I through Type III conflict.

Type IV conflict. Since it was not fully investigated in earlier work, the Type IV conflict is of particular interest here. Response mode percentages for Type IV conflict in the present study were quite similar to those for Type III, in most instances. Single (38.5% for Type III and 40.4% for Type IV) and compromise (33.3% and 28.8%) modes predominated for both conflict types, while detour, double, and disorganized responses were infrequent. Blocking was seen in a number of instances in response to both Type III (17.9%) and Type IV (16.6%) conflict. Hovland and Sears' (1938) percentages were quite different, with 72.5% blocking responses and 12.5% double responses to Type IV conflict.

Though resolution modes were similar in the present study, adequacy differentiated Type III and Type IV conflicts. A chi-square analysis of adequacy involved only those 25 Ss who differed in adequacy

from one conflict to the other. The resulting chi-square demonstrated that the Type IV conflict did result in a greater frequency of inadequate responses than did Type III, χ^2 (1) = 9.7, $p < .01$.

Response Lability

Mode lability analyses examined shifts in response mode from Trial 1 to Trial 2 and from Trial 2 to Trial 3 (Table 1). It was found that Type II conflict resulted in the greatest degree of lability, while Type I conflict and control trials showed the least total mode shifts, χ^2 (4) = 26.1, $p < .001$.

Lability data were further examined to determine the extent to which *S*s tended to shift toward or away from each major response mode with increased conflict experience. Summed over conflict types, there were significant differential tendencies to shift away from the single response mode over trials, χ^2 (1) = 4.25, $p < .05$, and toward the adequate double response mode, χ^2 (1) = 7.11, $p < .01$, and the inadequate blocking response, χ^2 (1) = 7.12, $p < .01$.

A further analysis of shift involved grouping the various responses to form the 2 classes originally defined by Hovland and Sears (1938). Class 1 contains responses that generally terminate at a goal, including double, single, and disorganized responses; while Class 2 contains only the blocking and compromise modes, which do not terminate at goals. The number of shifts from Class 1 to Class 2 modes (24) was considerably greater than the number of shifts (8) in the reverse direction, χ^2 (1) = 8.0, $p < .01$; and the number of interclass shifts (23) from Trial 1 to Trial 2 was significantly greater than the number of shifts (9) occurring from Trial 2 to Trial 3, χ^2 (1) = 6.1, $p < .02$. In addition, frequency of interclass shifts increased with increasing conflict difficulty, with the greatest number occurring on Type IV conflict, χ^2(4) = 19.6, $p < .001$.

Two additional factors influencing response lability were suggested by the work of Lewin (1935) and Hovland and Sears (1938). Lewin hypothesized that a single

TABLE 1

Response Lability (% Shifts) as a Function of Conflict Type

Conflict	Response mode shifts	
	Trial 1 to Trial 2	Trial 2 to Trial 3
Type I	15.4	9.6
Type II	48.1	38.5
Type III	32.7	25.0
Type IV	38.5	25.0
Nonconflict	11.5	17.3

Note. All percentages are based on 52 possible shifts (one per *S*) for each conflict type between each 2 trials.

response to a given goal in a Type I conflict situation weakens the valence of the selected goal, resulting in a shift to the opposite goal. This hypothesis was not confirmed in the present study, since only 6.5% of the *S*s shifted goals on either the second or third trial of the approach–approach conflict. An alternative hypothesis is that response lability is more a function of the reaction type or mode itself than of the field situation or conflict type (Hovland & Sears, 1938). This contention was not supported by present results, since the major response modes with which Hovland and Sears were concerned—single, double, and compromise—were associated with the very similar shifting values of 30.3%, 37.3%, and 28.1%, respectively.

A final consideration was the extent to which response lability is a function of the adequacy of the initial response. Response lability between Trials 1 and 3 was greater for inappropriate than for appropriate responses. Specifically, the mean percentage shift across the 6 inappropriate response modes (inappropriate detour, double, compromise, single, disorganized, and blocking responses) was 60.2%, while the mean for appropriate modes (appropriate double, single, and compromise responses) was 16.8%. An exception was the inappropriate blocking response, which was among the most stable responses, with only 20% shifts.

Discussion

Two basic questions studied in the present investigation concerning the Type IV conflict

related to its resolution under conditions of negative start and its relationship to other conflicts, particularly Type III. First, it was apparent that under the negative-start conditions of the present study, Type IV conflict, like Type III, elicited markedly less blocking than occurred in the work of Hovland and Sears (1938). Type III and Type IV conflicts were also similar to each other in relation to further response modes, speed, and error measures, while differing from the other conflicts. Only adequacy differentiated the 2 conflicts, with Type IV producing less adequate responding.

Both the Type III and Type IV conflicts, as well as the remaining 2 conflict types, showed some effects of repeated exposure. The overall increase in speed with practice was accompanied by differential mode lability, which was a function of conflict type and initial response adequacy. The hypotheses of both Lewin (1935), who stated that approach–approach conflict is subject to frequent shifting from goal to goal, and Hovland and Sears (1938), who said that lability is largely a function of initial response mode, had to be rejected on the basis of present results. The failure of the Lewin hypothesis may have been due to the use of left and right goals rather than, for example, forward and reverse goal points, since 93.9% of all single responses on Type I conflict were to the right-hand goal. Thus, right-lateral dominance, coupled with high preexperimental response strength for responses involving movement toward the right, may have affected results. Somewhat more massed practice may also be a necessary condition for the confirmation of the Lewin hypothesis. The negative-start condition of the present study, combined with the use of a maze rather than an open-field situation, and the fact that Hovland and Sears examined lability of only Type II conflict, may have contributed to the discrepant findings in the 2 studies concerning their hypothesis.

Further lability results showed that responses tend to shift away from inadequate early attempts and to shift from goal (Class 1) to nongoal (Class 2) modes, while interclass shifting becomes more frequent, the more difficult the conflict. These findings are related in that it is principally the Type III and Type IV conflicts that elicit inadequate responses, produce the most compromise and blocking (Class 2) responses, and have the highest interclass lability. In e[illegible] early inappropriate responses to Type III and IV conflicts often result in shifts to [illegible]r adequate compromise or inadequate blocking responses, depending on relevant individual difference variables.

The fact that approach–avoidance (Type II) was the most labile conflict when response class was not considered can be explained in relation to the Lewin (1935) and M[illegible] (1944) models. Both of these theories concluded that the response to Type II should reach some compromise point between the starting point and the conflicted goal. Lewin hypothesized, in addition, that *Ss* experienced in responding to approach–avoidance conflict situations may leave the field, a hypothesis consistent with the position that conflict produces a state of increased drive (Brown & Farber, 1951). In the latter case, leaving the field would be seen as one way of reducing the drive resulting from the conflict situation.

Present results for Type II showed single responses to the nonconflicted (51.9%) and conflicted (25%) goals to be most frequent, with compromise (3.2%) and detour (13.5%) modes relatively rare. The instructions and obvious boundaries of the paper maze in the present study made it difficult for *S* to literally leave the field. It was, however, quite possible to leave (detour or compromise) or avoid (single response to nonconflicted goal) the path leading directly to the conflicted goal and hence reduce the drive. Given these response alternatives, the unique ambiguity of the Type II conflict, resulting from the absence of any information (lights) concerning the status of the nonconflicted goal, may well have been responsible for the observed lability of this conflict. Moreover, the predominance of single responses is consistent with the fact that the single mode is the most practiced, due to its use on nonconflict trials.

In conclusion, given a conflict situation in which the starting point is negative and field departure difficult, the present study provided mixed support for several hypotheses drawn from earlier conflict theories and demonstrated that only response adequacy clearly differentiates Type IV conflict from Type III. Lability was shown to be determined primarily by initial response adequacy and conflict type. Further work systematically varying starting-point status, repetition of exposure, difficulty of field departure, situational ambiguity, and

relative goal location should make it possible to suggest some specific modifications of earlier conflict theories.

REFERENCES

Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables with suggestions toward a theory of frustration. *Psychological Bulletin*, 1951, **48**, 465–495.

Epstein, S., & Smith, B. Modes and adequacy of resolution of three basic types of cognitive-motor conflict. *Journal of Experimental Psychology*, 1967, **74**, 264–271.

Hovland, C. I., & Sears, R. R. Experiments on motor conflict. I. Types of conflict and their modes of resolution. *Journal of Experimental Psychology*, 1938, **23**, 477–493.

Lewin, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 1. New York: Ronald Press, 1944.

Scheffé, H. *Analysis of variance.* New York: Wiley, 1959.

Sears, R. R., & Hovland, C. I. Experiments on motor conflict: II. Determination of mode of resolution by comparative strength of conflicting responses. *Journal of Experimental Psychology*, 1941, **28**, 280–286.

Smith, B., & Epstein, S. Influence of incentive on adequacy and mode of conflict resolution. *Journal of Experimental Psychology*, 1967, **75**, 175–179.

Worell, L. Some ramifications of exposure to conflict. In B. A. Maher (Ed.), *Progress in experimental personality research.* Vol. 4. New York: Academic Press, 1967.

(Received March 17, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 56–61

MODES OF EXTRACTING INFORMATION IN CONCEPT ATTAINMENT AS A FUNCTION OF SELECTION VERSUS RECEPTION PARADIGMS

NEAL S. SMALLEY[1]

State University of New York College at Plattsburgh

Measures of 2 modes of inference were developed: the V measure counts inferences about attributes which are varied from a focus instance, and the C measure counts inferences about constant attributes. In 2 experiments Ss in a selection paradigm were compared with a yoked group in a reception paradigm. Both experiments supported the prediction that the reception group would give more C inferences, but found no difference in V inferences. The results are explained in terms of the selective attention to S's current hypothesis which is produced by the task of selecting, rather than receiving, information. The use of the V and C scores to measure focusing and scanning strategies is discussed.

The purpose of this investigation was to examine 2 methods of extracting information from instances in a concept attainment (CA) task. In particular, it was concerned with whether the method differs when *S* can choose the instances he wishes to examine (selection paradigm), as opposed to a situation where the choice of instances is not in his control (reception paradigm).

Attempts to experimentally compare the selection and reception paradigms (Hunt, 1965; Huttenlocher, 1962; Laughlin, 1972; Lowenkron & Johnson, 1968; Murray & Gregg, 1969; Schwartz, 1966) have based the comparison on effectiveness of performance as measured by number of trials (instances chosen) to criterion or solution time. No consistent pattern of results, favoring either selection or reception conditions, has emerged.

One alternative approach is to focus research attention on the more fundamental processes at each step of the CA task rather than on effectiveness of overall performance. The present investigation focused on the methods employed by *S*s to extract information from each instance encountered, employing measures of 2 such methods which were identified in a previous, unpublished study using the selection paradigm. For each instance which *S* chose he was asked for a reason for his choice and, after he was told whether the card was positive (an exemplar) or negative (a nonexemplar), he was asked for any inference he had drawn on the basis of the instance. The statements which *S*s made were classified into 3 types—V, C, and VC—which are illustrated in Table 1. The V inference is a statement about one or more attributes (or values of those attributes) which are varied from the positive example which *S* is given at the beginning of the trial. The C inference is a statement about one or more attributes (or values) which are constant, or the same as the attributes of the given instance. The VC inference is a statement about both varied and constant attributes. Table 1 gives examples of the 3 inference modes for each of the 4 main types of situations that commonly occur in conjunctive CA tasks. The task given as an example has stimuli with 4 attributes (represented as letters) and 2 values for each attribute (represented as numbers immediately following each letter). The concept shown is a 2-element conjunctive concept, and the inferences given are the logically correct ones for each of the 3 inference types. In addition to minor differences in wording, the inferences actually given by *S*s showed an occasional logical error and were often less complete in the cases where multiple inferences are possible (e.g., Situation I–C).

Employing a measure of mode of infor-

[1] Requests for reprints should be sent to Neal S. Smalley, Department of Psychology, State University College, Plattsburgh, New York 12901.

mation extraction has the additional advantage of overcoming a problem inherent in the design of experiments comparing reception and selection paradigms. In order to equate the groups on stimuli received, previous studies have employed the yoked group design in which each reception *S* received the sequence of instances actually chosen by an *S* in the selection group. However, Lowenkron and Johnson (1968), following Church (1963), have argued that *S*s differ in their ability to solve CA problems and that selection *S*s will choose sequences of instances whose difficulty is appropriate to their abilities. Since reception *S*s are paired with selection *S*s at random, the correlation between ability level and sequence difficulty will not be present in the reception group, and this may result in a negative bias against the reception group.

A study which uses as its dependent variable a measure of information extraction can avoid this bias because differences between *S*s in their abilities to extract information should be minimal. With a 2-element conjunctive concept, the task of drawing an inference from each instance is relatively easy and almost all *S*s should produce a logically correct inference for each instance, as was the case in the exploratory study. Thus *S*s should not differ substantially in their ability to draw inferences but rather in the manner in which they draw them.

The present study compared selection and reception groups, hypothesizing that *S*s in the selection condition would make more V inferences and that *S*s in the reception condition would make more C inferences. Consistent with previous suggestions (Dienes & Jeeves, 1965; Hunt, 1965), the selection condition, which provides the opportunity for manipulation of the type and order of information to be examined, is assumed to thereby permit *S*s to take a more active role, systematically varying the attributes of the stimulus materials and thus focusing their attention on V attributes. On the other hand, the reception condition, since it does not provide this opportunity, is assumed to engender a more passive role in which *S* attends to patterns of relationships common to the instances observed and thus focuses on C attributes.

TABLE 1

THREE INFERENCE TYPES FOR THE FOUR MOST FREQUENT SITUATIONS WITH A TWO-ELEMENT CONJUNCTIVE CONCEPT

Situation type	Instance	Category	Inference mode	Inference
I	A1,B1,C1,D1	+		(The given instance.)
	A2,B1,C1,D1	+	V	A1 is not part of the concept.
			C	(B1 and C1), (B1 and D1), and (C1 and D1) are possible concepts.
			VC	(A1 and B1), (A1 and C1), and (A1 and D1) are not the concept.
II	A2,B1,C1,D1	−	V	A1 is part of the concept.
			C	(B1 and C1), (B1 and D1), and (C1 and D1) are not the concept.
			VC	(A1 and B1), (A1 and C1), and (A1 and D1) are possible concepts.
III	A2,B2,C1,D1	+	V	Neither A1 nor B1 is part of the concept.
			C	(C1 and D1) is the concept.
			VC	(A1 and B1), (A1 and C1), (A1 and D1), (B1 and C1), and (B1 and D1) are not the concept.
IV	A2,B2,C1,D1	−	V	Either A1 or B1 or both are part of the concept.
			C	(C1 and D1) is not the concept.
			VC	(A1 and B1), (A1 and C1), (A1 and D1), (B1 and C1), and (B1 and D1) are possible concepts.

Note. The V inference is a statement about one or more attributes (or values of those attributes) which are varied from the positive example which *S* is given at the beginning of a trial. The C inference is a statement about one or more attributes (or values) which are constant, or the same as the attributes of the given instance. The VC inference is a statement about both varied and constant attributes.

EXPERIMENT I

Method

Subjects and design. The *S*s were 28 college freshmen who volunteered during a summer orientation program prior to their first year of college. Two groups were employed, a reception (REC) and a selection (SEL) group. In order to control the stimuli received by the 2 groups, a yoked group design was used in which *S*s in the REC and SEL groups were randomly paired and each REC *S* received the instances chosen by his paired SEL *S*. Thus *S*s were randomly assigned to groups with the necessary restriction that a particular *S* not be assigned to the REC group unless an unpaired SEL *S* who had already completed the experiment was available.

Materials. The instances were hand-lettered on 5.1 × 7.6 cm. white cards and arranged in an ordered

array. Each instance contained 4 capital letters—A, H, Q, and R—each of which had either the number 1 or the number 2 printed as a subscript below it. The letters were the attributes and the subscripts were the values of those attributes; therefore this was a 4-attribute, 2-value task. The array contained all possible instances ordered so that each card was adjoined on all sides only by cards which varied from it in a single attribute.

Since the purpose of the experiment was to study types of inferences, it was desirable to control possible sources of error and confusion for the *Ss*. Exploratory studies seemed to indicate that unitary stimuli such as the usual geometric figures used in CA experiments occasionally produced such confusion, presumably due to failures to separate the attributes combined in the single geometric figures. The letter-and-subscript form of the stimuli in the present experiment was an attempt to reduce this confusion by physically separating the attribute value pairs.

Procedure. The directions read to *S* explained the construction of the cards, stated that for each problem the concept would involve only 2 attributes conjunctively combined, and gave an example of one such concept. The directions for the 2 groups differed only in whether the *Ss* were instructed to select cards from the array (the SEL group) or were told that *E* would give them the cards (REC group).

All *Ss* received 4 problems, or different concepts, each of which began with the card $A_1H_1Q_1R_1$, which was always positive. For each successive card, 4 steps were followed. First, a SEL *S* chose his card; a REC *S* was given the card chosen by his paired SEL *S*. Second, *Ss* in both groups were asked: "Do you have any hypotheses about what the concept might involve?" Third, the category of the card (positive or negative) was announced, and the card was placed in either of 2 columns headed with a [illegible] "Does that card lead you to any particular conclusions?" In the SEL group a problem was completed when *S* announced the correct concept—unless that announcement was a guess, that is, unless *S* had not chosen enough cards to be logically certain. When such guesses occurred, *S* was encouraged to pick additional cards to "test" his guess. In the REC group a problem was not completed until *S* had received all of the cards which his paired SEL *S* had chosen for that problem, even if the REC *S* attained the concept earlier than the SEL *S*. If a REC *S* had not attained the concept when the SEL *S*'s cards ran out, he was encouraged to go back over the cards and attempt to figure out the concept. He was not given any additional cards. These procedures for the 2 groups insured that each pair of *Ss* would have the opportunity to examine exactly the same cards from which to draw inferences.

The sessions were tape-recorded using a cassette recorder which was in full view of *S*. In addition, *E* wrote down each card that *S* chose or was given and attempted to note every card to which *S* pointed [illegible] his verbal explanations.

These procedures differ from some versions of the CA task in that the cards [illegible] removed from the array and placed in posi[illegible]nd negative colum[illegible] where they remained i[illegible] view during the entire problem. This was done [illegible]imize the effects of differences in memory ab[illegible] and to avoid the confusion which err[illegible] memory for past in[illegible]stances can sometimes pro[illegible]

Results

Each statement [illegible]n answer to the questions was first [illegible]orized as an inference or a nonin[illegible] An inference was defined as a sta[illegible] which included one or more attribu[illegible]e letters A, H, Q, R) or values (a l[illegible]umber combination such as A_1) and [illegible]elated these to the concept in some [illegible]amples of such relationships are st[illegible] that the attributes or values a[illegible]ot, or possibly [illegible] the concept [illegible] Only inferences we[illegible] further.

Each inference [illegible]ced in one of [illegible] categories: V, C, or VC. An inference was placed in the V categ[illegible]ll the attributes or attribute–value [illegible] mentioned were ones which were different from, or varied from, the given card. Since the given card was always $A_1H_1Q_1R_1$, the V inferences were always about attributes which had a subscript of "2" on the card being discussed. An inference was put in the C category if all the attributes or attribute-value pairs mentioned were the same as, or constant with, the given card. Thus C inferences were always about attributes which had a subscript of "1" on the card in question. Finally, an inference was classed as VC if the attributes or attribute-value pairs mentioned in the inference included at least one which was V type and one which was C type, as just defined. Examples of each type of inference in the situations most common for this task are given in Table 1. In cases such as the statement for Situation 1, each of the possible concepts which *S* actually gave was counted as a separate C inference.

There were, then, 3 response measures: the total number of V, of C, and of VC responses over the 4 problems. The SEL and REC groups were compared on each of these 3 measures by means of a matched

groups *t* test. The mean number of C inferences in the REC group was 24.6, and in the SEL group it was 14.9. The difference was in the predicted direction and was statistically significant, t (13) = 3.15, $p < .01$. The V means 10.8 (SEL) and 8.9 (REC) were in the predicted direction but were not significant, t (13) = .68. No prediction was made about the VC measure, nor was it found to significantly distinguish the 2 groups, t (13) = 1.40.

Thus the results were consistent with the hypothesis that the reception paradigm tends to produce a more passive, pattern-oriented approach to extracting information from instances than does the selection approach. However, there was no support for the hypothesis that the selection paradigm produces a more active, systematic approach to information extraction.

For comparison with previous studies employing measures of effectiveness of performance, the number of trials (instances) required to attain each of the 4 concepts was computed. A concept was said to be attained on the first trial in which (a) S announced the correct concept without mentioning any alternative concepts, and (b) it was logically possible to definitively conclude what the concept was. Using this criterion, the mean number of trials required to solve all 4 problems was 16.7 in the REC group and 17.4 in the SEL group, a difference which was statistically significant, t (13) = 2.86, $p < .02$. The superior performance of the REC group obviously contradicts the possibility of a negative bias against the reception group resulting from the yoked control design. Additional evidence against this bias was also furnished by the error rates, which failed to suggest a difference in difficulty between the stimulus sequences for the 2 groups. There were no differences in either the total percentage of logically erroneous inferences (5.2% for SEL and 4.8% for REC) or in errors in particular types of inferences: for V inferences, 6.0% (SEL) and 4.0% (REC), and for C inferences, 3.4% (SEL) and 4.6% (REC).

One possible objection to Experiment I centers on the part of the procedure that asked *S* to give any hypotheses he held about the concept. In answering this question, which was asked just after he had made his choice of an instance, the SEL *S* may have given the hypothesis he was attempting to test with that choice, rather than giving all the possible hypotheses which might be tested with that instance. The REC *Ss*, unrestricted by a prior hypothesis, may have been able to give more complete lists of possible hypotheses. Since the hypotheses were combined with the inferences given in response to the question asking for *S*'s conclusions, it is possible that the observed differences were due primarily to the difference between the number of hypotheses given. However, a repetition of the t tests based only on the inferences given to the "conclusions" question produced findings nearly identical to those obtained initially: for the C measure, t (13) = 3.90, $p < .001$; for V, t (13) = .92, *ns*; and for VC, t (13) = 1.45, *ns*.

This revision in the C measure does not, however, entirely meet the objection. The "hypotheses" question itself may have directed the REC *Ss* to look for possible hypotheses which they would not otherwise have produced. Therefore, asking for hypotheses may have facilitated the performance of REC *Ss* while giving no corresponding help to SEL *Ss*, who were presumably set by the hypothesis they had already produced in the process of selecting an instance. To eliminate this interpretation of the results, a second experiment was run which replicated the first one except that the request for hypotheses was omitted from the procedure.

Experiment II

[illegible] tion and a yoked reception group.

Procedure. All aspects of the procedure [illegible] student who was the [illegible]

the concept might involve?" was omitted from the procedures for both groups. Thus the only opportunity to make verbal inferences was in response to the question asking for conclusions from each card.

Results

The 3 measures were derived in exactly the same way as in Experiment I. Again, there was a considerable difference between the mean number of C inferences in the REC group (24.6) and the SEL group (17.8), a difference which was statistically significant, t (15) = 3.51, $p < .01$. For the V measure, neither the direction nor the magnitude was as predicted—mean SEL = 5.6, mean REC = 6.8, t (15) = .62. Again, the VC measure did not produce a significant difference, t (15) = .62. In this experiment, there was no significant difference in trials to criterion—means of 19.6 (SEL) and 19.9 (REC), t (15) = .50—providing less strong but still consistent evidence contrary to the hypothesis of negative bias against the REC group. As in Experiment I, no significant differences in error rates were found. For the SEL and REC groups respectively, the percentages of erroneous inferences were 4.3% and 7.2% for all inferences combined, 3.3% and 3.7% for V inferences, and 3.9% and 8.4% for C inferences.

The C score may be seen not only as a measure of a mode of drawing inferences but also as a measure of the completeness with which these inferences are drawn. Recall that in Situations I and II, 1, 2, or all 3 possible concepts may be given. Each concept given was counted into the total C score. Thus, the more such concepts S gave for each instance, the more complete was the inference process and the greater was the information extracted from the instance. However, we may ask whether the difference in C scores simply reflects greater completeness in the REC condition than in the SEL condition, or whether there is an actual preference in the REC condition for inferences of the C type. To get some indication of this from the present data, the C score was revised by counting only one C inference per instance in cases where 2 or 3 such inferences were given; the mean C scores for the REC and SEL groups then became 17.5 and 15.2, respectively. The difference is in the predicted direction, although not significant, t (15) = 2.01.

Discussion

The major difference between the selection and reception conditions appears to be in the completeness of the C inferences drawn by Ss in the selection condition. Thus, instead of the hypothesis of active and passive roles initially suggested, a better formulation can now be offered. The task of selecting instances may set S to extract only information relevant to his current hypothesis and may therefore result in his ignoring additional information which the instances contain. When S receives rather than selects instances, he may be less set by any hypotheses he has and therefore able to extract more information from each instance.

More generally, the modes of information extraction identified in this study appear to underlie the CA strategies described by Bruner, Goodnow, and Austin (1956). In *focusing* strategies, S selects an instance which varies one or more attributes from the given instance. His goal is to gain information about those attributes, and so his inferences should be about varied (V) attributes. In *scanning* strategies, S tests one or more possible concepts which are directly represented in the test instance. Since these possible concepts will also be embodied on the given card, the inferences drawn will be about constant (C) attributes. However, the V and C scores are not perfect indexes of the strategies employed because Ss' verbal statements include additional deductions about attributes and concepts not directly tested by a particular card. The V and C scoring system compares favorably with the measure of focusing developed by Laughlin (1965) in that it requires fewer, simpler decisions by the scorer and is applicable to both reception and selection paradigms.

Perhaps because the V and C measures reflect modes of extracting information from instances rather than either patterns of instance selection or effectiveness of performance, they may prove applicable to a variety of experimental situations and problems.

References

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking*. New York: Wiley, 1956.

Church, R. M. The varied effects of punishment on behavior. *Psychological Review*, 1963, **70**, 369–402.

Dienes, Z. P., & Jeeves, M. A. *Thinking in structures.* London: Hutchinson Educational, 1965.

Hunt, E. Selection and reception conditions in grammar and concept learning. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 211–215.

Huttenlocher, J. Effects of manipulation of attributes on efficiency of concept formation. *Psychological Reports*, 1962, **10**, 503–509.

Laughlin, P. R. Selection strategies in concept attainment as a function of number of persons and stimulus display. *Journal of Experimental Psychology*, 1965, **70**, 323–327.

Laughlin, P. R. Selection versus reception concept-attainment paradigms for individuals and cooperative pairs. *Journal of Educational Psychology*, 1972, **63**, 116–122.

Lowenkron, B., & Johnson, P. J. Yoked group comparisons of selection and reception paradigms of concept attainment: A biased procedure. *Psychological Reports*, 1968, **23**, 1143–1149.

Murray, F. S., & Gregg, R. E. Reception versus selection procedures in concept learning. *Journal of Experimental Psychology*, 1969, **82**, 571–572.

Schwartz, S. H. Trial-by-trial analysis of processes in simple and disjunctive concept-attainment tasks. *Journal of Experimental Psychology*, 1966, **72**, 456–465.

(Received March 22, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 62–66

EFFECT OF NOISE ON THE STROOP TEST

L. R. HARTLEY[1] AND R. G. ADAMS

Medical Research Council, Applied Psychology Unit, Cambridge, England

Noninterference tests and 2 versions of the Stroop color interference test were used to test *S*s in loud noise and quiet. Interference and noninterference tests were performed during the first and last 10 min. of a 30-min. exposure to noise and quiet. Interference in noise and quiet was measured by taking the difference between performance on interference and noninterference tests. The *S*s in Experiment I who were tested at both exposure durations showed increased interference in noise. In Experiment II, exposure duration and practice were assessed independently. The brief exposure to noise was beneficial and decreased interference, and the long exposure increased interference, suggesting a cumulative adverse effect of noise.

Exposure to loud noise undoubtedly produces changes in human performance efficiency (Broadbent, 1953). These performance changes have been attributed to the amount of stimulation being received: excessive stimulation from intense noise is supposed to overarouse *S* and cause impairment (Broadbent, 1963), but a moderate amount of stimulation can cause some improvement (Davies & Hockey, 1966; McGrath, 1963). These performance changes have occurred toward the end of a lengthy (30-min.) test in continuous loud noise (Broadbent & Gregory, 1963). The changes occurring with continued exposure to loud noise include increases in errors and pauses in a serial reaction task (Wilkinson, 1963) and increased attention to high-probability sources of information (Hockey, 1970). Broadbent (1971) suggests that one interpretation of these results is that the increased arousal accompanying exposure to intense noise leads to an impairment of the mechanism that filters information from the environment for further processing. Some of the evidence for this view derives from a series of studies involving the visual identification of common and uncommon words in noise. It was found that the uncommon word was more difficult to see in noise, but only when both classes of words were present. Noise did not alter the ratio of misperceptions of common and uncom[illegible] words. Noise clearly affected the [illegible] tion of the stimulus rather than cha[illegible] the bias in responding to one or the [illegible] class of words.

In the Stroop test (Jensen & Roh[illegible] 1966; Stroop, 1935) color names written in inks of different hues, excluding ink of the hue the name indicates, are presented to *S*s, who are required to name the hue of the ink and not the color name. The comparison or control task may involve naming monochromatic color names or naming the hue of the ink in which nonverbal symbols are written. Responding to the hue of color names takes longer than either of the comparison tasks because both name and hue are appropriate to the same set of responses. Hence, the irrelevant feature of the stimulus (color name) must be excluded from the analysis. The color name dimension must be filtered from the hue dimension (Dyer, 1971).

It would be expected that if filtering is affected by noise-induced arousal, the Stroop test should show impairment in noise which would increase with length of exposure to noise, as other tests have shown it does. In the following experiments, versions of the test not requiring an overt verbal response were used.

Experiment I

Method

Subjects. Eighteen enlisted men served as *S*s, and each *S* was tested individually. No *S* had

---

[1] Requests for reprints should be sent to L. R. Hartley, Medical Research Council, Applied Psychology Unit, 15 Chaucer Road, Cambridge, England.

done the test before. All *S*s had normal hearing; less than 35-db. loss in one ear or 30-db. loss in both ears at any of 7 test frequencies covering the audible range.

Materials. Two packs of 100 cards, one experimental (E) and the other control (C), were used. Cards in the E pack displayed 5 color names red, blue, green, violet, and black—written in uppercase lettering in inks of the 5 hues. No color name was ever written in the hue the name indicated. There were 5 cards of each hue/name combination in the pack. Each card in the C pack displayed 5 crosses drawn in any one of the 5 hues of the E pack. There were 20 cards of each hue in the pack. The *S*s sorted each of the 2 packs as quickly and accurately as possible into 5 piles corresponding to each of the hues.

Procedure. All 18 *S*s were tested in conditions of noise (N) and quiet (Q). These 2 tests were given on separate days at between 8:00 and 9:00 A.M. Each *S* sorted both the E and C packs in the first and last 10 min. of a 30-min. exposure to N and Q. The *S*s read magazines in the interval between the tests at the beginning and end of the exposure.

Broad-band noise was used in all conditions, having constant energy per cycle between 50 and 4,000 Hz. In N, this was presented at 100 db., and in Q, at 70 db., measured on the C scale, having a substantially flat frequency response.

Half of the *S*s had the N condition on their first test and half had the Q condition. Presentation order of the E and C packs to each group of 9 *S*s was counterbalanced within each *S*. Prior to the first test, *S*s practiced sorting both packs of cards in Q. Each *S* was told his scores after sorting each pack, and the scores were also displayed on a group chart near *S*.

TABLE 1

MEAN TIME (IN MIN.) TO SORT EACH CARD PACK UNDER EACH CONDITION

Time of sorting task	Card-pack stimuli	
	Color names	Crosses
Quiet condition		
First 10 min.	2.6156	2.3167
Last 10 min.	2.2630	2.0722
Noise condition		
First 10 min.	2.6990	2.2593
Last 10 min.	2.3880	2.0620

Note. Time of sorting task refers to its occurrence within the 30-min. exposure to either quiet or noise.

Results

The times taken to sort each pack were analyzed using a repeated measures analysis of variance with logarithmic transformation of the scores to achieve normality of distribution. Interactions of each factor with *S*s were used as error terms.

Sorting on the second test, following the 20 min. of prior exposure, was faster than sorting on the first test in every condition ($p < .008$). Pooling tests in the N and Q conditions, the type of pack sorted interacted with the position of the test in the $\frac{1}{2}$-hr. exposure, $F\ (1, 17) = 6.32$, $p < .025$. The difference in time taken to sort names and crosses was less at the end of the exposure than at the beginning. Comparing tests in the N and Q conditions independently, N increased the time taken to sort the E pack by about 3% and decreased the time taken to sort the C pack by about 1% (Table 1).

The difference in time taken to sort E and C packs was nearly twice as large in N (22.97 sec.) as in Q (14.69 sec.), $F\ (1, 17) = 4.55$, $p < .05$. Although noise caused an increase in the difference in sorting time of the 2 packs, there was no evidence that it interacted with the position of the test in the $\frac{1}{2}$-hr. exposure ($p = .25$). The increased difference in sorting time of the 2 packs can therefore be regarded as similar at the beginning and at the end of $\frac{1}{2}$ hr. of noise.

EXPERIMENT II

Experiment I examined whether noise increased the interference of color name with the selection by hue and whether this interference increased with the duration of noise exposure. The fact that there was no increased interference with a long as compared to a short exposure to noise could be due to a number of factors; in particular, the fact that duration of exposure to noise was confounded with practice at test, and the latter reduced the interference. In Experiment II, these 2 variables are assessed independently and a modified version of the test, involving a much smaller motor component, was used.

Method

Subjects. Two groups of 16 *S*s each were tested. All were enlisted men, none of whom had taken part in Experiment I. The *S*s were tested in groups of up to 6 at a time. All *S*s had normal hearing; less than 35-db. loss in one or 30-db. loss in both ears at any of 7 test frequencies covering the audible range.

Materials. In the modified version of the Stroop test, the material was presented in a written form, and *S* ticked the appropriate item, rather than sorting it from a set of cards. The material was presented in a series of sheets, each containing 32 lines of material. On E sheets, each line was made up of the color names, red, green, brown, black, and blue, written in inks of these hues. No color name was written in the hue that the name indicated. Each line on the sheets contained 1 stimulus color name on the left and the 5 possible response color names on the right. Each *S* was instructed to select from the response set the color name appropriate to the hue of the stimulus name. As a control, a noninterference (C) test was also used. The format of these C sheets was identical to that of the E sheets, and they contained the same set and number of color names as the E sheets; however, color names were all printed in black ink. Each *S* was instructed to select from the response set the same color name as the stimulus color name.[2] In both C and E tests, *S* crossed out the response word in as many lines of material as he could in 5 min.

Procedure. All *S*s were tested twice, once in both the N and Q conditions. A practice test was also given. These 3 tests took place on 3 consecutive days between 8:00 and 9:15 A.M.

Broad-band noise having equal energy per octave was used in all conditions. This noise was presented at 95 db. in the N condition and at 70 db. in the Q condition, measured on the C scale. The differences in sound-pressure level and spectrum of the noise between the 2 experiments reported here were necessitated by administrative reasons outside *E*'s control.

One group of 16 *S*s (Group 1) was exposed to only 10 min. of Q or N, performing both C and E tests immediately following entry into the experimental sound room: a second group of 16 *S*s (Group 2) performed both C and E tests in the last 10 min. of a 30-min. exposure to N and Q. The *S*s in the latter group read magazines prior to the tests during the first 20 min. of both the N and Q conditions. In each condition, *S* did both E and C tests for the 5-min. period allowed. Half of the *S*s in each group were tested in N on their first test day and half in Q. On each test day half of the *S*s had the E test first and half had the C test first. Both tests were scored in terms of number of lines responded to correctly. All [illegible] were instructed before each test to work [illegible] quickly and accurately as possible, completing [illegible] lines of the test as they could. The *S*'s score [illegible] displayed in a group chart on the wall near him [illegible]

Results

The number of lines completed [illegible] analyzed using analysis of variance [illegible] Experiment I, and Wilcoxon's test. [illegible]he *S*s worked faster on their second test than on their first on E sheets ($p < .001$) and on C sheets ($p < .02$), Wilcoxon's test. Every *S* worked more slowly on the [illegible] than the C test in every condition. As Table 2 shows, there was a tendency for *S*s to work slower on the C sheets and faster on E sheets in the 10-min. exposure to N, as compared to Q. On the other hand, the group with 30-min. exposure tended to work slower on E sheets and faster on C sheets in N as compared to Q. None of these differences were reliable however. The changes in speed were about 7% on E sheets and about 1% on C sheets.

The differences in number of items correctly completed on E and C tests in each condition were used to assess the interference caused by N as compared to Q. The *S*s made very few errors on either test. The average number of errors on each C and E test was less than 1. Taking the difference, or interference score, between C and E tests, there was no overall effect of N as compared to Q. There was however, a strong interaction between Interference and Duration of Exposure to N and Q, $F\ (1, 30) = 15.81$, $p < .001$. This interaction occurred because the interference, or difference between C and E tests, was less in the N condition (45.13) than in the Q condition (55.50) in the first 10-min. exposure, $F\ (1, 30) = 7.46$, $p < .025$. Only 2 out of 16 *S*s did not show less interference in N. However, considering the other group of *S*s, there was more interference with the 30-min. exposure to N (57.82) than in Q (40.00), $F\ (1, 30) = 8.38$, $p < .01$. Again, only

[2] This monochromatic control was selected in preference to material containing stimuli of different colors for ease of duplication and preparation, with a view to possible large-scale use of the test in the future. This experiment was part of a program of evaluation of the test for use in field studies of environmental stress.

2 out of 16 *S*s did not show greater interference in the 30-min. exposure to N than to Q. These 2 effects of noise did not relate systematically to either the order in which N and Q conditions were received or the order in which the 2 tests were received. Neither could there be any effect of instructional bias upon the group since not all members of a group had C and E tests presented in the same order.

TABLE 2

Mean Number of Lines Completed under Each Cindition

Group	Stimulus sheet	
	Control	Experimental
	Quiet condition	
1	170.13	114.63
2	172.44	132.44
	Noise condition	
1	169.38	124.25
2	176.88	125.06

Note. Group 1 *S*s were exposed to only 10 min. of either quiet or noise; Group 2 *S*s were tested after 20 min. of a 30-min. exposure to quiet or noise.

Discussion

The second version of the Stroop test is clearly more sensitive than the card-sorting version. In both experiments, nearly all the change in interference was contributed by performance on the experimental rather than the control material. The *S*s in Experiment II showed a much more reliable change in interference than those in Experiment I, notwithstanding the higher sound-pressure level in the first experiment. This is probably due in part to the much smaller motor and larger perceptual component in the test in Experiment II.

In addition, in the first experiment, the response of sorting by color was unrelated to the semantic description of the hue. In Experiment II, although *S* did not verbalize, he selected the appropriate semantic response in relation to the hue. Pritchatt (1968) considered 2 conditions similar to these experiments and also obtained considerably more interference when *S* was required to press a key labeled by color name rather than hue. The difference in interference between the 2 versions is generally consistent with the view of Hock and Egeth (1970) that much of the interfering effect is due to disruption of the encoding of the name of the hue by the verbal dimension of the stimulus or its aftereffect.

The effect of noise on this interference depends on duration of exposure; interference was reduced by the short exposure, confirming the results of O'Malley and Poplawsky (1971). The increased impairment during the long exposure is consistent with the change in performance in other tests during a long exposure to noise, where impairment has increased with exposure. Hartley (1973) has found that a 40-min. exposure to noise causes impairment in a serial reaction test performed only in the last 20 min. of the exposure. This impairment was worse than when 20 min. of noise was presented in the last half of 40 min. of performance. These results indicated that continuous noise led to a progressive increase in impairment, independent of task duration. In Experiment II, the difference in performance between the brief and the long exposure may also be construed as further evidence that the effect of noise on performance changes with exposure. Impairment appeared following the long exposure without any prior performance.

There is some evidence to connect the difference in interference between short and long exposure to noise with changes in arousal. Agnew and Agnew (1963) and Tecce and Happ (1964), using threat of electric shock, and Callaway (1959), using an amphetamine, obtained results indicating reduced interference under these conditions. On the other hand, Callaway (1959), using amylobarbital, and Ostfeld and Aruguete (1962), using hyoscine, obtained results indicating increased interference under these conditions. It would be reasonable to assume, although without substantiating evidence, that stimulants and shock raised arousal, while the sedatives depressed arousal. The beneficial and detrimental effects of short and long exposure to noise may likewise be connected with changes in arousal level.

A significant feature of Experiment II is the absence of any auditory cues in performance of the test. Some auditory cues, either generated by *S* or by the apparatus, have been present in nearly all tests that have shown impairment in noise. Although the fact

that noise has shown its effect at the end of the test suggests that masking of these auditory cues did not contribute to the noise impairment observed, this factor could not be ruled out (Kryter, 1970). For example an interaction between time spent on the task and loss of auditory information could account for impairment in noise appearing at the end of the test. The results of Experiment II, a task with no auditory component, clearly indicate that the adverse effect of noise need not be due to the loss of these cues, particularly in view of the initial beneficial effect of noise also obtained. Nevertheless, these results do not exclude the possibility that loss of task-independent auditory information is connected with impairment when the loss is prolonged for 20 min.

REFERENCES

Agnew, N., & Agnew, M. Drive level effects on tasks of narrow and broad attention. *Quarterly Journal of Experimental Psychology*, 1963, **15**, 58-62.

Broadbent, D. E. Noise, paced performance and vigilance tasks. *British Journal of Psychology*, 1953, **44**, 295-303.

Broadbent, D. E. Possibilities and difficulties in the concept of arousal. In D. N. Buckner & J. J. McGrath (Eds.), *Viligance: A symposium.* New York: McGraw-Hill, 1963.

Broadbent, D. E. *Decision and stress.* London: Academic Press, 1971.

Broadbent, D. E., & Gregory, M. Vigilance considered as a statistical decision. *British Journal of Psychology*, 1963, **54**, 309–323.

Callaway, E. The influence of amobarbital (amylobarbitone) and methamphetamine on the focus of attention. *Journal of Mental Science*, 1959, **105**, 382–392.

Davies, D. R., & Hockey, G. R. J. The effects of noise and doubling the signal frequency on individual differences in visual vigilance performance. *British Journal of Psychology*, 1966, **57**, 381–389.

Dyer, F. N. The duration of word meaning responses: Stroop interference for different preexposure of the word. *Psychonomic Science*, 1971, **25**, 229-231.

Hartley, L. R. Effect of prior noise or prior performance on serial reaction. *Journal of Experimental Psychology*, 1973, **101**, 255–261.

Hock, H. S., & Egeth, H. Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology*, 1970, **83**, 299-303.

Hockey, G. R. J. Signal probability and spatial location as possible bases for increased selectivity in noise. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 37–42.

Jensen, A. R., & Rohwer, W. D., Jr. The Stroop Colour-Word test: A review. *Acta Psychologica*, 1966, **25**, 36–93.

Kryter, K. D. *The effects of noise on man.* New York: Academic Press, 1970.

McGrath, J. J. Irrelevant stimulation and vigilance performance. In D. N. Buckner & J. J. McGrath (Eds.), *Vigilance: A symposium.* New York: McGraw-Hill, 1963.

O'Malley, J. J., & Poplawsky, A. Noise induced arousal and breadth of attention. *Perceptual and Motor Skills*, 1971, **33**, 887-890.

Ostfeld, A. M., & Aruguete, A. Central nervous system effects of hyoscine in man. *Journal of Pharmacology and Experimental Therapeutics*, 1962, **137**, 133–139.

Pritchatt, D. An investigation into some of the underlying associative verbal processes of the Stroop colour effect. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 351–359.

Stroop, J. R. Studies of interference in serial verbal reaction. *Journal of Experimental Psychology*, 1935, **18**, 643–662.

Tecce, J. J., & Happ, S. J. Effects of shock-arousal on a card sorting test of colour word interference. *Perceptual and Motor Skills*, 1964, **19**, 905-906.

Wilkinson, R. T. Interaction of noise, with knowledge of results and sleep-deprivation. *Journal of Experimental Psychology*, 1963, **66**, 332-337.

(Received March 24, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 67–80

VARIABILITY AND CONTROL IN DICHOTIC MEMORY [1]

STANLEY R. PARKINSON [2]

Arizona State University

Experiments are reported in which *S*s were instructed on each trial to attend to one of 2 stimultaneously presented messages. Following a procedure developed by M. P. Bryden, order of report was manipulated by requiring *S*s to either recall the attended message first and unattended message second (AU report), or the unattended message first and attended message second (UA report). While the group data of Experiments I and II provided a clear replication of Bryden's results, extensive individual differences in unattended-ear recall functions were obtained. Conventional single-channel digit spans were assessed for *S*s in Experiment III and *S*s were grouped according to their digit span. Correlations between digit span and unattended-ear performance yielded coefficients of .88 and .84 in AU and UA reports, respectively. The theoretical implications posed by the relation between digit span and dichotic memory and the increase in experimental precision which is gained by the knowledge of digit span warrant the further use of digit span as a concomitant variable in dichotic memory research.

Dichotic memory refers to a set of experimental operations in which *S*s are given 2 different lists of items simultaneously, one to each ear, with instructions to reproduce as many of the items as they can following list presentation (Broadbent, 1971).

In the special case of the dichotic experiment where stimulus presentation is rapid (1½–2 pair/sec) and where both lists consist of a single stimulus type (e.g., digits) recorded in the same voice, *S*s tend to report items from one ear and then from the other ear (ear-by-ear report) rather than reporting items by pair (pair-by-pair report) according to their order of arrival (Broadbent, 1954; Bryden, 1962). When forced to recall pair by pair under these stimulus conditions, accuracy levels are decreased.

The reliability of these findings motivated Broadbent (1971) to formulate a model for dichotic memory performance. Broadbent proposed the existence of 2 systems: an "S" or storage system capable of parallel processing and thus of holding information received by 2 ears simultaneously, followed by a "P" or perceptual system limited to sequential information processing. According to the model, auditory–verbal information enters the organism via the S system. A selective filter, operating in time between the S and P systems, selects the message from one ear for entry into the P system and attenuates processing of the message received by the other ear. The P system processes (categorizes) the selected ear message and that message is reported first during recall. The initially unselected message is held in the S system until processing of the first message has been completed, at which time it gains access to the P system. Another way of stating the operation of the model is that an individual attends to and rehearses the message received by one ear (P system) while the message received by the other ear is held in a very-short-term buffer or echoic store (S system). Following recall of the attended message, *S* shifts his attention to and recalls the second message.

As Broadbent (1971) has noted, the S and P system nomenclature is somewhat misleading as the S system is viewed as a parallel processing system while the P system is assumed to be sequential in operation. Due to this confusion, and to

[1] This research was supported in part by the U.S. Air Force Human Resources Laboratory, monitored by the Flying Training Division under Contract F-41609-72-C-0037. The author wishes to express his gratitude to Lora Hubbard for her assistance in all phases of this research.

[2] Requests for reprints should be sent to Stanley R. Parkinson, Department of Psychology, Arizona State University, Tempe, Arizona 85281.

more closely align the terminology with other memory models, the term "echoic store" (Neisser, 1967) will be used when referring to a very-short-term auditory buffer in which information is relatively unanalyzed (S system), and "active store" (Kroll, Parkinson, & Parks, 1972; Sternberg, 1969) will be used when referring to a system consisting of attention and rehearsal processes (P system).

A finding of major importance for the present purpose is that recall accuracy for the ear reported second in dichotic memory tasks is generally lower than that for the ear reported first (e.g., Bryden, 1964). In terms of the model under consideration this finding implies that the echoic store is (*a*) limited in duration, and/or (*b*) limited in capacity, and/or (*c*) vulnerable to either input or output interference (or both) resulting from first-ear processing.

Broadbent (1957) conducted an experiment in which the 2 messages of dichotic pairs differed in length. The message to the ear to be reported first consisted of 6 digits, while 2 digits were presented to the ear to be reported second. In order to test a decay or limited duration hypothesis, the 2-digit message was presented simultaneously with either the first 2, middle 2, or final 2 digits of the 6-digit message. Presentation rate was 2 pair/sec, making the difference in time between the extreme conditions on the order of 2 sec. Broadbent found more accurate recall when the 2-digit message coincided in presentation with the final 2 digits of the 6-digit message, a result which is consistent with the hypothesis that echoic storage is subject to rapid decay (1–2 sec.).

A similar time estimate was offered by Murray and Hitchcock (1969). They instructed *S*s to either silently code (without observable mouth movement) or mouth code (articulate silently) one of a pair of 5-digit messages presented dichotically. In a third group, labeled the "not code group," *S*s were instructed to say *the* as each pair of stimuli arrived. At the end of each stimulus list, *S*s were probed for their retention of a single digit on either the "coded" or "uncoded" message. Murray and Hitchcock found that, even with a probe recall procedure which minimizes response interference, only one item was recalled above chance levels in the uncoded message of the silently-code and mouth-code groups and in both messages of the not-code group. Given the stimulus presentation rates employed (1 and 2 pair/sec), they estimated that the duration of echoic storage was between 1 and 2 sec.

In another series of experiments directed towards determining the temporal parameters of echoic storage, however, Bryden (1971) obtained results which suggest that the time course of echoic storage is considerably longer (several seconds) than previously indicated. Using a dichotic memory procedure, Bryden instructed his *S*s to "attend" to one ear (either left or right) during list presentation. Order of report was manipulated by instructing the *S*s to either recall attended ear first and unattended ear second (attended–unattended or AU order), or to recall unattended ear first and attended ear second (UA order). Bryden argued that if unattended ear items were retained in an echoic store of very brief duration (1–2 sec.), recall should be impaired when they were reported second (AU) relative to a condition in which they were reported first (UA).

The results which Bryden (1971) obtained were not in accord with his hypothesis; i.e., the accuracy of unattended-ear recall did not vary as a function of order of report. Positively increasing serial position functions were found with unattended-ear recall regardless of order of report. Serial position functions on the attended ear were relatively flat with both report orders, but unlike the unattended ear, attended-ear recall was impaired when reported second. This latter result would be expected if it were assumed that attended-ear items were retained in a rehearsal-dependent active store, as reporting of unattended items in UA reports would prevent rehearsal of attended-ear items.

Bryden (1971) attempted to explain his failure to detect a difference between unattended-ear recall as a function of order of report (and to retain the notion of an echoic store) by making 3 assumptions: (*a*) The S system or echoic memory lasted longer than the delay of reporting U items in AU reports; thus, no impairment of unattended items occurred. (*b*) Echoic storage is less vulnerable to output interference than is active storage. This follows from his finding that the difference between attended reports ($A_1 - A_2$) was greater than the difference between unattended reports ($U_1 - U_2$). (*c*) The echoic store is of limited capacity (1–2 items). The latter interpretation was derived from the positively increasing serial position functions obtained with unattended-ear recall.

The present series of studies represents an attempt to further study the nature of recall processes in dichotic memory. The first experiment reported here was a replication of the Bryden (1971) procedure, in which attention and recall instructions were manipulated in a within-*S*s design. In the second experiment objective measures of attention instructions were obtained by requiring *S*s to either write down or verbalize the attended-ear message as it was presented. The third experiment was concerned primarily with providing a means for accounting for some of the variability in dichotic memory performance by grouping *S*s according to conventional digit spans.

Experiment I

Method

Subjects. Twelve students enrolled in an introductory experimental psychology course at Arizona State University served as *Ss* in the present experiment. The *Ss* were all right-handed and varied in age from 19 to 25 yr.

Materials and apparatus. Subjects were tested individually in a sound reduction chamber. Subjects received 42 dichotic lists each consisting of 2 different lists of 4 digits presented simultaneously, one to each ear. Both channels (left and right) were recorded by the same male speaker at the rate of 2 pair/sec. Stimulus lists were recorded on a TEAC model TCA 42 4-channel tape deck and they were played to *S* through stereo headphones (Superex Pro-BV). The stimulus lists were constructed from the set of numbers 0–9 and no number appeared twice on a given trial.

Procedure. Each *S* was instructed to attend exclusively to one ear and to ignore the other ear during stimulus presentation, even though messages to both ears were to be recalled. One half of the *Ss* were told to attend to the right ear and one half were told to attend to the left ear. The ear of attention remained constant for each *S* during the course of the experiment.

Subjects received 2 blocks of 21 trials, each block consisting of 5 practice trials followed by 16 test trials. During the first block of trials, 6 *Ss* (3 attending to their right ear and 3 to their left ear) were instructed to recall the digits in the attended ear first and then to recall the digits in the unattended ear (AU order), while the remaining 6 *Ss* (again 3 right ear and 3 left ear) were given instructions to recall first the digits in the unattended ear followed by the digits in the attended ear (UA order). All *Ss* were reversed in the second block of 21 trials: those reporting AU in the first block reported UA in the second while those reporting UA initially were told to report AU in the second block.

Serial recall instructions were given in all conditions; i.e., in AU reports *Ss* were instructed to recall the attended-ear digits in the order in which they were received and then to recall the unattended-ear digits again in the order received. While *Ss* were instructed to recall as many of the digits as they could remember on each trial, no attempt was made to force *Ss* to recall all 8 digits. If an *S* remembered only 2 unattended-ear digits, he was encouraged to report only those 2 in the order in which they were presented. Consider an example in which the unattended message was 6, 8, 4, 2. If the *S* remembered only 4 and 2, he was instructed to report only 4 and 2. It was thought that if *Ss* were encouraged to guess, and thereby to report more items, the additional response interference might cause them to forget the items which they did remember. For the same reason, in the previous example, *Ss* were not instructed to say ***blank, blank***, 4, 2, even though that type of report would have given a more accurate indication of serial position.

No attempt was made in this experiment to objectively determine if *Ss* were in fact following the attention instructions (i.e., if *Ss* were attending to only one ear during stimulus presentation). Order of report instructions, however, was closely monitored. If *Ss* reported in reverse order (e.g., AU in a UA condition), the tape was stopped and *Ss* were reminded that order of report was important. This procedure proved to be successful, as reversals of this type occurred in less than 1% of the total observations.

Results

While *S*s were instructed to recall as many of the digits as they could in the

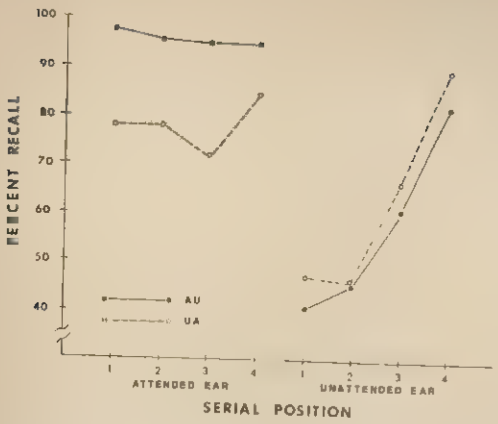


FIGURE 1. Group data, Experiment I: percentage correct recall as a function of attention (attended–unattended ear), order of report (attended–unattended, AU; unattended–attended, UA) and serial position.

order in which they were received (serial recall), incomplete reports occurred in the majority of trials. As there is no adequate way to deal with omissions in serial recall, a free-recall scoring procedure was used. Items were scored as correct in the serial position in which they were presented if they were recalled in the correct ear. Consider the following examples in which the attended message was 4, 5, 9, 1, and the unattended message was 2, 6, 3, 8. If, given an AU report instruction, *S* reported 4, 5, 9, 1, 2, 6, 3, 8, he received a perfect score on both his attended and unattended ears. If an *S* reported 4, 5, 9, 1, 3, 8, he received a perfect attended-ear score and 3 and 8 were scored as correct in the third and fourth serial positions, respectively, on the unattended ear. If an *S* reported 4, 9, 5, 1, 2, 8, he still received a perfect attended-ear score, even though 9 and 5 were transposed, and 2 and 8 were scored as correct responses in the first and fourth serial positions, respectively, on the unattended ear.

Raw scores were converted to percentage correct recall and a 2 Between × 3 Within analysis of variance was computed. In this analysis, the ear of attention (left or right) and the sequence of report instructions (AU–UA or UA–AU) were between factors while report instructions (AU and UA), attention (A and U), and serial position (1, 2, 3, 4) were within-groups facto[r]

The results obtained in this experime[nt] (see Figure 1) were remarkably similar [to] those obtained by Bryden (1971, Figure [5]). Neither of the between factors ear of [at]tention or sequence of report instructions produced reliable differences; however, ea[r] of attention (right ear superiority) approached significance, F (1, 8) = 2.5[?], $p = .149$, $MS_{\text{Error}} = 1{,}283.86$.

Performance on the attended [ear] w[as] superior to performance on the una[ttended] ear, F (1, 8) = 27.47, $p < .001$, $MS_{[\text{Error}]}$ = 1,282.78, and as in the Bryden (19[71]) study there was a significant interaction b[e]tween attention and serial position, F (3, 24) = 13.43, $p < .001$, $MS_{\text{Error}} = 322.41$. The serial position functions for the attended ear, whether reported first or second, were relatively flat, while positively increasin[g] serial position functions were obtained fo[r] the unattended ear. The Order of Repo[rt] × Attention interaction was significan[t], F (1, 8) = 26.65, $p < .001$, $MS_{\text{E}[\text{rror}]}$ = 230.91, indicating a greater differen[ce] between attended-ear scores ($A_1 - A_2$) than between unattended-ear scores ($U_1 - U_2$).

The group data are fully consistent with the findings of Bryden (1971). The failure to detect a difference in unattended-ear recall as a function of order of report is compatible with the hypothesis that echoic storage lasts for several seconds. The positively increasing serial position functions obtained with unattended reports suggests that the echoic store is limited in capacity.

However, while the group data were supportive of the Bryden (1971) findings and interpretations, there was extensive between-*S*s variability in the present experiment. The individual functions of all 12 *S*s are shown in Figures 2 and 3. With regard to the unattended ear, it can be seen that the 6 *S*s in Figure 2 have individual functions which reflect the group performance shown in Figure 1. These 6 *S*s all show unattended-ear performance to be a positively increasing function of serial position. The 6 *S*s shown in Fig-

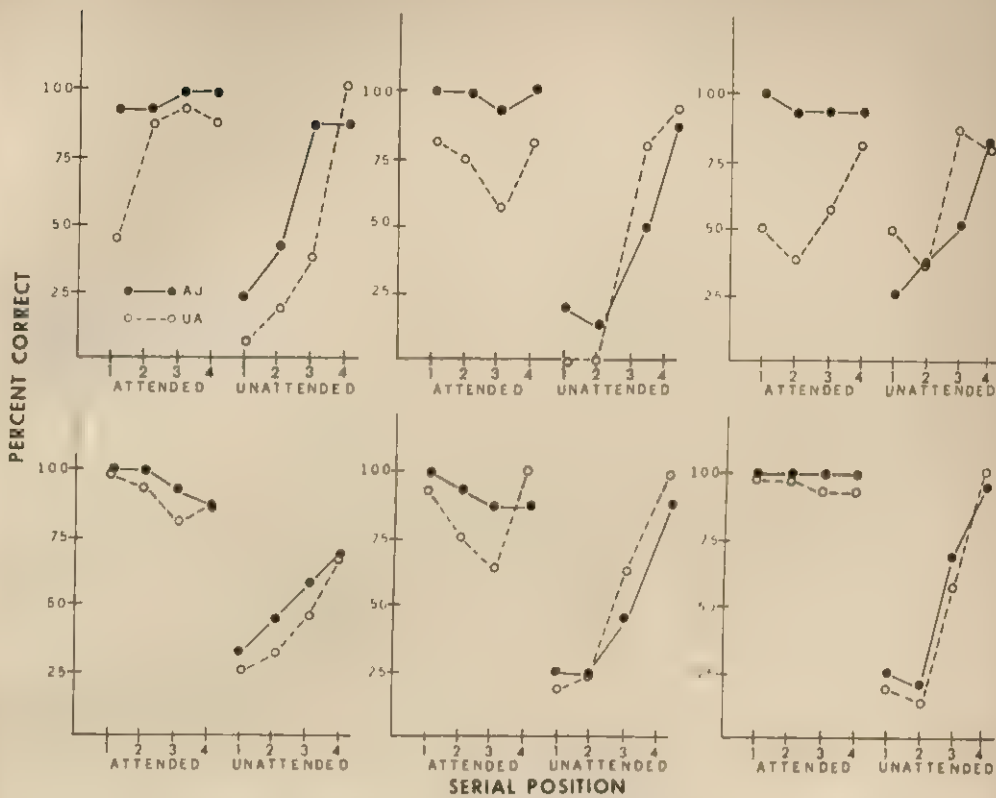


FIGURE 2. Individual data, Experiment I: percentage correct recall as a function of attention (attended unattended ear), order of report (attended unattended, AU, unattended attended, UA), and serial position.

ure 3, however, show individual unattended-ear functions which are widely divergent from the group average.

In an attempt to determine the source of variability between *S*s, a more detailed analysis of the response patterns was made. In the AU order of report, all 12 *S*s responded with 4 digits from the attended ear before switching to the unattended ear. This response pattern was obtained in 190 of the 192 trials.

In contrast to the consistency found between *S*s in AU reports, there was considerable variability in UA reports. Five *S*s responded with an average of 2 or fewer U items in UA reports, while the remaining 7 *S*s responded with an average of 4 U responses. All of the 5 *S*s responding with 2 or fewer U responses showed positively increasing serial position functions on the unattended ear, while 6 of the 7 *S*s responding with 4 U items showed a pronounced absence of that effect. Whether or not recall accuracy on the unattended ear is a positively increasing function of serial position is dependent on the number of items reported from that ear.

The variability in unattended-ear performance in this experiment precludes any definitive theoretical statement about the properties of echoic storage. What experimental operations or designs might be employed in order to reduce or to account for this variability?

EXPERIMENT II

In the second experiment *S*s were again instructed to attend to only one ear (balanced left or right over *S*s) during stimulus presentation. In order to insure more objectively that *S*s were following the instructions, *S*s were told to either

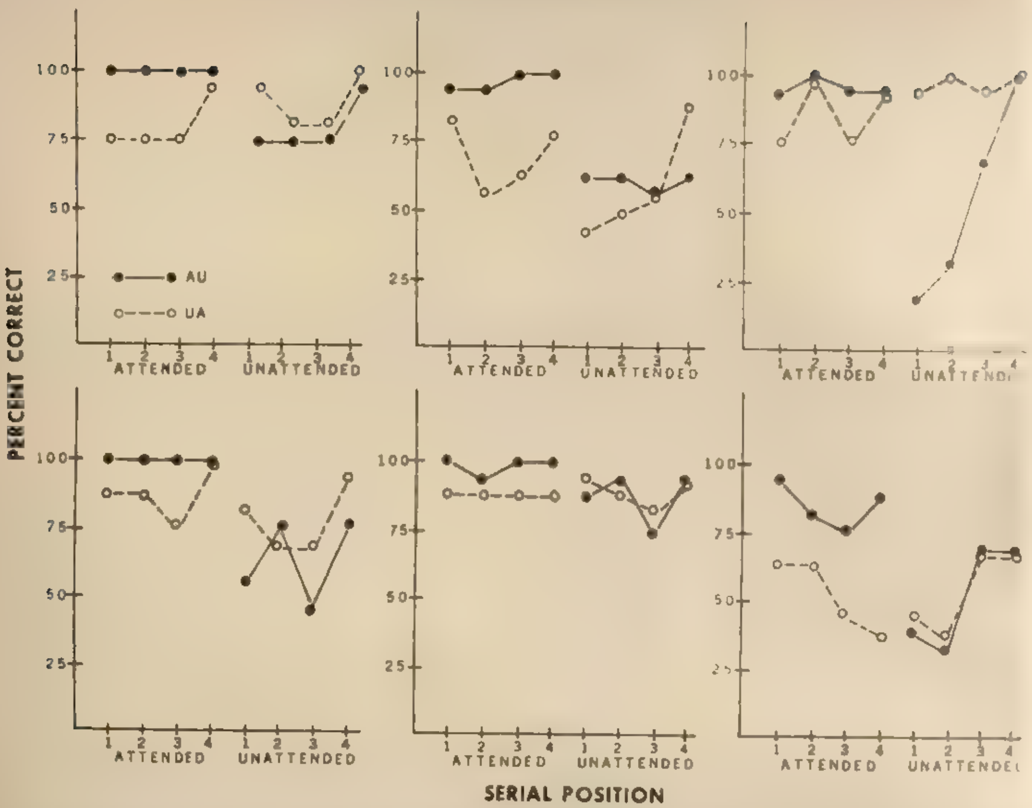


FIGURE 3. Individual data, Experiment I: percentage correct recall as a function of attention (attended–unattended ear), order of report (attended–unattended, AU; unattended–attended, UA), and serial position.

write down or verbalize the message received by the attended ear as it was presented.

If failure to follow the attention instructions contributes to the variability found in this type of task, then in Experiment II a disparity in recall functions should exist between conditions in which *S*s follow instructions (attended message perfectly monitored during stimulus presentation) and when they do not (imperfect monitor of attended-ear message).

Another reason for conducting the second experiment concerns attended–unattended performance as a function of type of monitor. Both the overt acts of verbalizing and writing a message during presentation constitute types of rehearsal. Given that attended-ear performance is mediated by an active storage system, comprised of attention and rehearsal processes, both monitor types should result in uniformly high attended-ear performance. If unattended-ear performance is mediated by an echoic store in which information is relatively unanalyzed, subsequent auditory stimulation might be expected to interfere with the maintenance of and/or recall from this system. A monitor which produces further auditory stimulation by way of auditory–verbal feedback (verbal monitor) might, therefore, result in more interference than a monitor which does not contribute auditory stimulation (written).

Method

Subjects and apparatus. Twenty-four students enrolled in an introductory experimental psychology class served as *S*s in this experiment. The materials and apparatus employed were identical to those of the first experiment.

Procedure. Subjects were randomly assigned to one of 2 groups: verbal monitor and written monitor. Each *S* was told to attend to only one ear during stimulus presentation. One half of the *S*s

TABLE 1

PERCENTAGE CORRECT RECALL, EXPERIMENT II, FOR WRITTEN AND VERBAL GROUPS, WITH AND WITHOUT PERFECT MONITORING

Group	Attended		Unattended	
	Perfect	With errors	Perfect	With errors
Written	90.81	88.74	67.28	65.66
Verbal	93.62	91.40	43.00	39.30

in each group were instructed to attend to the right ear and one half were told to attend to the left ear. As in the first experiment the ear of attention for each S remained constant during the course of the experiment.

Subjects received the same 2 blocks of 21 trials as those used in the first experiment. The sequence of instructions for recall, AU–UA and UA–AU, was balanced for both groups.

Written monitor. Subjects were instructed to write down the attended-ear message as it was presented. During this task S inserted his right arm through an aperture in a wooden stand. A piece of black opaque plastic around the aperture prevented S from seeing his written response. While S was writing his attended-ear response, E observed both the S and the VU meter in the preamplifier to insure that the written monitor was synchronized with the presentation of the message. After a few practice trials Ss started in the 2 blocks of 21 experimental trials.

Verbal monitor. Subjects were told to verbally monitor the attended-ear message. The E wrote down the verbal monitor as well as the verbal recall response for the attended and unattended ears. Again E observed the VU meter in an attempt to insure synchrony of verbal monitor and stimulus presentation.

Results

The first question posed with regard to this experiment asked if failure to follow instructions (shown by an imperfect monitor) were a factor contributing to performance variability. If it were, a disparity in recall functions should be observed when Ss followed instructions (perfect monitor) and when they did not (imperfect monitor). These scores (collapsed over serial position) were obtained for each individual, and the group means are shown in Table 1. Correlated t tests showed all of these differences to be unreliable, all ps $> .05$. All trials with perfect and imperfect monitors were included in further analyses.

As in the first experiment, a free-recall scoring analysis was used. Raw scores were converted to percentage correct recall and a 3 Between $\times$ 3 Within analysis of variance was conducted. This analysis was the same as that in the first experiment, with the addition of type of monitor as a third between-Ss variable.

The results shown in Figures 4 and 5, depicting performance for the written and verbal monitor groups, respectively, indicate that as in the first experiment (*a*) attended-ear recall was superior to unattended-ear recall, F (1, 16) = 112.60, $p < .001$, MS_{Error} = 1,131.48; (*b*) the Attended–Unattended Ear $\times$ Serial Position interaction was significant, F (3, 48) = 76.94, $p < .001$, MS_{Error} = 130.66, with attended-ear serial position functions flat and unattended-ear functions positively increasing; and (*c*) the Order of Report (A–U vs. U–A) $\times$ Attention (A vs. U) interaction was reliable, showing a greater difference between attended-ear reports than between unattended-ear reports, F (1, 16) = 20.95, $p < .001$, MS_{Error} = 141.94.

Of more interest in the present results are the effects of type of monitor on various aspects of recall patterns. While the written monitor group was superior to the verbal monitor group, F (1, 16) = 18.70, $p < .001$, MS_{Error} = 553.29, the difference was due solely to the unattended ear. The interaction (shown in Figure 6)

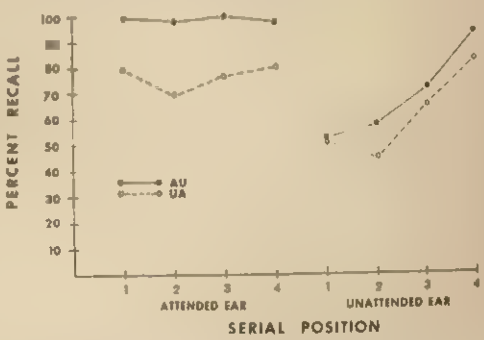


FIGURE 4. Written monitor group, Experiment II: percentage correct recall as a function of attention (attended–unattended ear), order of report (attended–unattended, AU; unattended–attended, UA), and serial position.

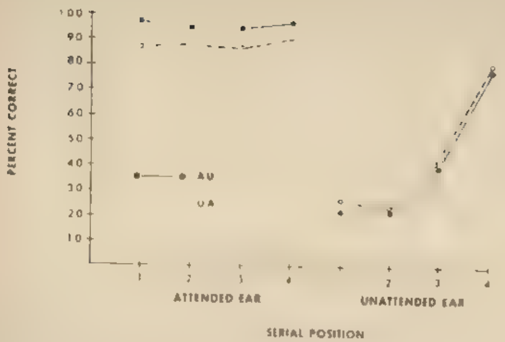


FIGURE 5. Verbal monitor group, Experiment II: percentage correct recall as a function of attention (attended–unattended ear), order of report (attended–unattended, AU; unattended–attended, UA), and serial position.

between type of monitor and ear of report was reliable, F (1, 16) = 17.49, $p < .001$, $MS_{\text{Error}} = 1{,}131.48$, as was the triple order interaction between type of monitor, ear of report, and serial position, F (3, 48) = 4.41, $p < .008$, $MS_{\text{Error}} = 130.66$.

These results, showing both differential recall functions for attended and unattended ear and selective impairment of unattended-ear recall with a verbal monitor, are compatible with a dual storage model in which attended-ear performance is mediated by an active store and unattended-ear performance is mediated by an echoic store. While interesting, however, the results of this experiment did not contribute to understanding of the mecha-

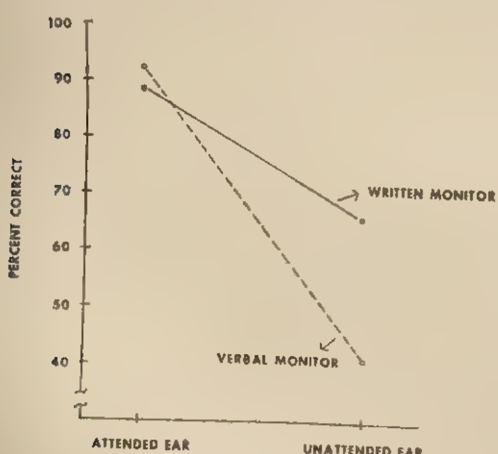


FIGURE 6. Percentage correct recall as a function of type of monitor and attention.

nisms underlying variability in dicho memory performance.

On reexamination of the individual d shown in Figures 2 and 3, it was obvi that some *Ss* were recalling more ite than were others. Perhaps this dispari in number of items recalled was a functi of capacity, with variability in dichot memory performance being determined l individual differences in "immediate mem ory span" (Miller, 1956).

In order to test the capacity hypothesi digit spans were assessed for each *S* Experiment III and *Ss* were grouped cording to digit span.

EXPERIMENT III

Method

Tests of digit span were administered to potential *Ss* entering the laboratory. The stin for measuring spans consisted of 5 sets of each set being comprised of 2 lists of each len ranging from 3 to 15 digits. Each *S* was te individually and received all 5 list sets. Stim presentation was binaural and presentation was set at 2 digit/sec. During each span assess ment, *Ss* received an ascending series of lists with instructions to reproduce each list in the corr t order immediately following its termination. This procedure continued until an *S* erred on both lists of a given length, at which time his digit span was assigned as one less than that number (e.g., if an *S* missed both lists with 8 digits, his digit span for that list set was 7). Each *S* was assigned an average digit span following 5 such assessments. In order to be included as an *S* in this experiment, certain criteria had to be met. The 5 digit spans for an *S* had to exhibit a clear mode; 3 of the 5 span estimates had to be identical. In addition, no single span estimate could vary by more than one unit from the mode. Thus, while scores of 6, 5, 6, 7, 6 resulted in a digit span of 6, a set of scores such as 5, 7, 6, 7, 7 resulted in exclusion from the experiment, as 5 was 2 units less than the mode of 7. This criterion is rather strict and it resulted in the exclusion of many individuals, but in a first approximation towards determining the effects of digit span in dichotic memory performance, a strict criterion was thought to be advisable.

It seemed essential to cover a wide range of spans, so the design was set for groups with span estimates of 5, 6, 7, 8, 9, and 10 digits. The attempt to get 8 *Ss* per group was successful for spans of 6, 7, 8, and 9; however, only 4 *Ss* were obtained for the 5- and 10-digit groups. Therefore, a total of 40 *Ss* was included in this experiment.

Subjects within all digit-span groups were balanced with regard to both ear of attention (left

or right) and order of report sequence (AU–UA, UA–AU).

The instructions and experimental procedure were identical to those of the first experiment with the following exceptions: (*a*) each *S* received 10 practice trials prior to a test block, and (*b*) each test sequence included 30 trials. Therefore, in this experiment each *S* received 20 practice ists and 60 test lists.

Results

The results for all *S*s are shown (*a*) collapsed over digit-span groups for both AU and UA reports in Figure 7, and (*b*) as a function of digit-span group for AU and UA reports in Figures 8 and 9, respectively. While all of the digit-span groups are included in the figures, in order to retain equal group size only the groups comprised of 8 *S*s each (i.e., Digit-Span Groups 6, 7, 8, and 9) were included in the analysis of variance. It is obvious from Figures 8 and 9 that any difference detected between digit-span groups would be accentuated by the addition of Groups 5 and 10.

Examination of Figure 7 reveals results comparable to those obtained in Experiments I and II. Indeed, as in the other

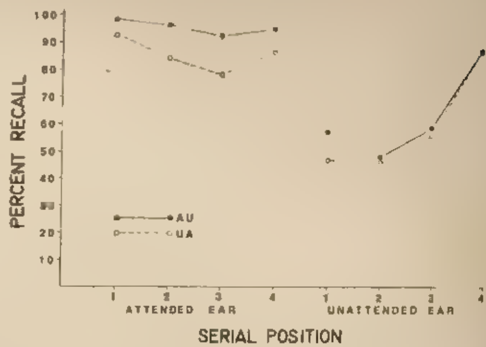


FIGURE 7. Group data, Experiment III: percentage correct recall as a function of attention (attended–unattended ear), order of report (attended–unattended, AU; unattended-attended, UA), and serial position.

experiments the following results were reliable: (*a*) the AU order produced higher recall than the UA order, F (1, 16) = 24.52, $p < .001$, $MS_{\text{Error}} = 245.01$; (*b*) attended ear was superior to unattended ear, F (1, 16) = 258.90, $p < .001$, $MS_{\text{Error}} = 434.48$; (*c*) terminal serial positions were more accurately recalled than initial serial positions, F (3, 48) = 55.43, $p < .001$, $MS_{\text{Error}} = 196.66$, a result which is due primarily to the positively increasing

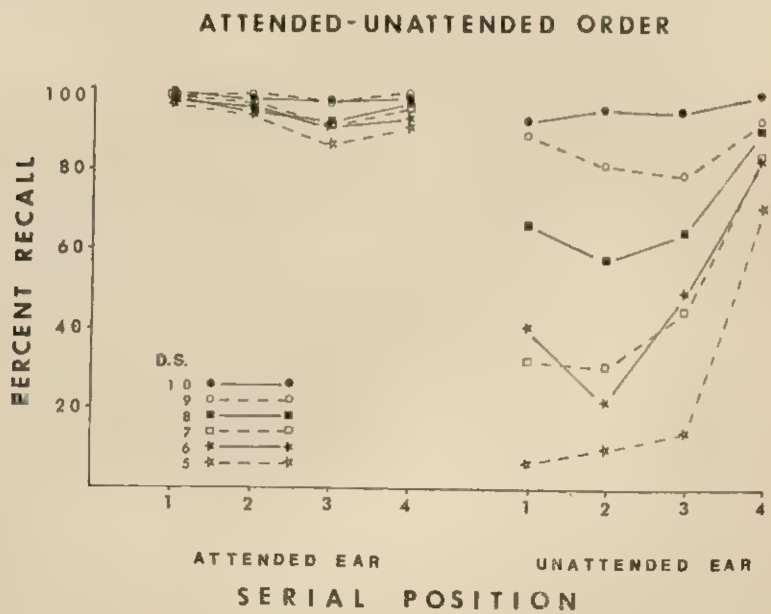


FIGURE 8. Percentage correct recall with the attended–unattended report order as a function of digit span (DS), attention (attended–unattended ear), and serial position.

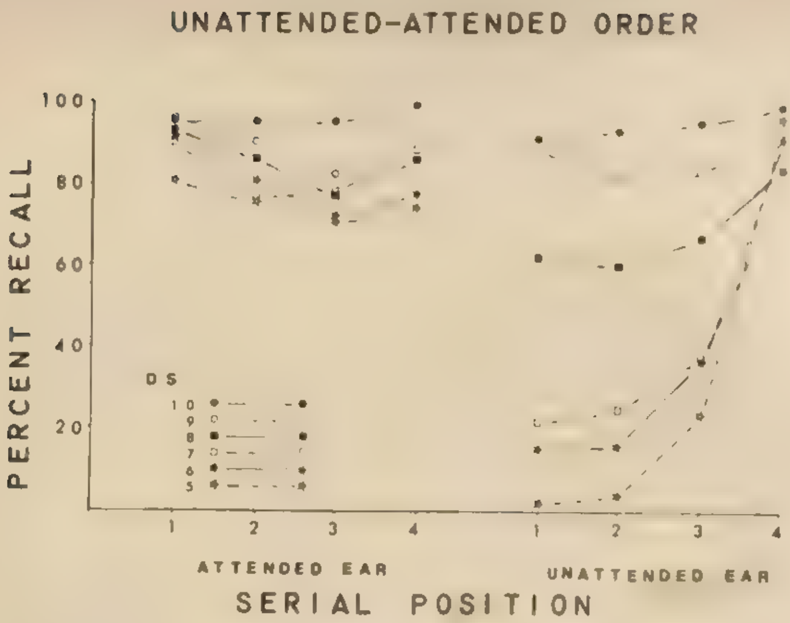


FIGURE 9. Percentage correct recall with the unattended-attended report order as a function of digit span (DS), attention (attended-unattended ear), and serial position.

serial position functions in the unattended ear, F (3, 48) = 58.18, $p < .001$, MS_{Error} = 193.51; and (d) there was a greater disparity between attended-ear scores when reported first and second than between unattended-ear reports as a function of report order, F (1, 16) = 7.45, $p < .01$, MS_{Error} = 215.60.

The effect of digit-span groups as a between-*Ss* variable is highly reliable, F (3, 16) = 23.24, $p < .001$, MS_{Error} = 742.74; and while there is a small effect of digit span on the attended ear when reported second, the most substantial effect of digit span is on unattended-ear performance. This was confirmed in statistical analysis by a Digit Span × Attended–Unattended Ear interaction F (3, 16) = 23.88, $p < .001$, MS_{Error} = 434.48, a Digit Span × Serial Position interaction, F (9, 48) = 5.76, $p < .001$, MS_{Error} = 196.66, and a Digit Span × Attention × Serial Position interaction, F (9, 48) = 7.21, $p < .001$, MS_{Error} = 193.51.

TABLE 2

COEFFICIENTS OF CORRELATION AND DETERMINATION BETWEEN ATTENDED-EAR AND UNATTENDED-EAR RECALL ACCURACY AND DIGIT SPAN

Reports	Attended ear		Unattended ear	
	r	r^2	r	r^2
Attended–unattended	.474	.225	.881	.776
Unattended–attended	.625	.391	.844	.712

While reliable, the interactions do not provide us with any indication of the strength of the relationship between digit span and dichotic memory performance. The total number of items correct in the dichotic task was calculated for each S (N = 40), and the correlation between digit span and dichotic performance was determined. This analysis yielded a very high correlation coefficient, r = .891. The coefficient of determination (Hays, 1963), r^2, gives an indication of the strength of the linear relationship between 2 measures. The coefficient in the present experiment indicates that 79.39% of the variance in dichotic memory can be accounted for by a linear rule and digit span. In Table 2 separate coefficients of correlation and determination are given for each measure of dichotic performance, i.e., for attended- and unattended-ear recall in both AU and

UA reports. These figures indicate that most of the variance accounted for by digit span is in unattended-ear recall.

In the first experiment, the nature of the unattended-ear serial position functions was shown to be due in part to differential report strategies. Positively increasing serial position functions were obtained when the mean number of unattended items reported was 2 (Figure 2), while more variable functions were obtained when an average of 4 unattended items was reported (Figure 3). The *S*s in Figure 2 were therefore making errors of omission, while the *S*s in Figure 3 were making primarily errors of commission, either in the form of interlist or extralist intrusions. An error analysis for the *S*s in the present experiment is provided in Table 3. In this table the number of errors is shown as a function of digit-span group and attention (attended–unattended ear). Errors were divided into 3 types: omissions—items presented on either the attended or unattended ear which were not reported during recall; interlist intrusions—items presented on the attended ear which were recalled as unattended items and items presented on the unattended ear which were reported as attended ear items; extralist intrusions—items reported during recall which were not members of the stimulus set.

As shown in Table 3, while low-span *S*s made more attended-ear errors than did high-span *S*s, no apparent differences were evident in the pattern of errors made by *S*s in the various digit-span groups. However, on the unattended ear there was an orderly trend of type of error as a function of digit span. Approximately 90% of the unattended errors made by *S*s with digit spans of 5, 6, and 7 were errors of omission, i.e., items were simply not reported during recall. This tendency was still prominent for *S*s with a digit span of 8, as 74% of the errors were omissions. Between Spans 8 and 10, however, there was a shift in the type of errors, with a reduction in errors of omission and an increase in errors of commission, so that by Span 10 the majority of errors were interlist intrusions.

TABLE 3

TYPE OF ERROR (OMISSIONS, INTERLIST INTRUSIONS, AND EXTRALIST INTRUSIONS) AS A FUNCTION OF ATTENTION AND DIGIT SPAN, EXPERIMENT III

Digit span	Omissions		Interlist intrusions		Extralist intrusions	
	No.	%	No.	%	No.	%
Attended ear						
5 ($n = 4$)	134	44.37	156	51.66	12	3.97
6 ($n = 8$)	213	50.84	193	46.06	13	3.10
7 ($n = 8$)	167	47.72	170	48.57	13	3.71
8 ($n = 8$)	116	42.96	135	50.00	19	7.04
9 ($n = 8$)	73	42.44	86	50.00	13	7.56
10 ($n = 4$)	12	29.27	23	56.10	6	14.63
Unattended ear						
5 ($n = 4$)	576	89.44	59	9.16	9	1.4
6 ($n = 8$)	932	91.46	78	7.66	10	.98
7 ($n = 8$)	904	89.33	93	9.19	15	1.48
8 ($n = 8$)	482	74.04	150	23.04	19	2.92
9 ($n = 8$)	219	54.89	154	38.59	26	6.52
10 ($n = 4$)	31	40.79	38	50.00	7	9.21

This analysis raises the interesting problem that response strategy (which is correlated with digit span) and not digit-span ability per se might be the factor which produced the results of these experiments. Due to the large number of incomplete reports (errors of omission), I was forced to use a free-recall scoring procedure; and as high-span *S*s responded with more unattended-ear items than low-span *S*s, their recall functions might have been inflated simply due to guessing. Two questions then arise: (*a*) What would the recall patterns of high-span *S*s look like if a scoring procedure (serial recall scoring) less sensitive to guessing were employed? and (*b*) What would happen if low-span *S*s were coerced to recall 8 digits per trial? With regard to the second question, would the recall patterns of low-span *S*s become more like those shown typically by high-span *S*s (which would mean that the results obtained were primarily a function of response strategy and not digit-span ability per se), or would they remain much the same as when these *S*s recalled in a manner of their own choice (which would be more indicative of digit-span ability)?

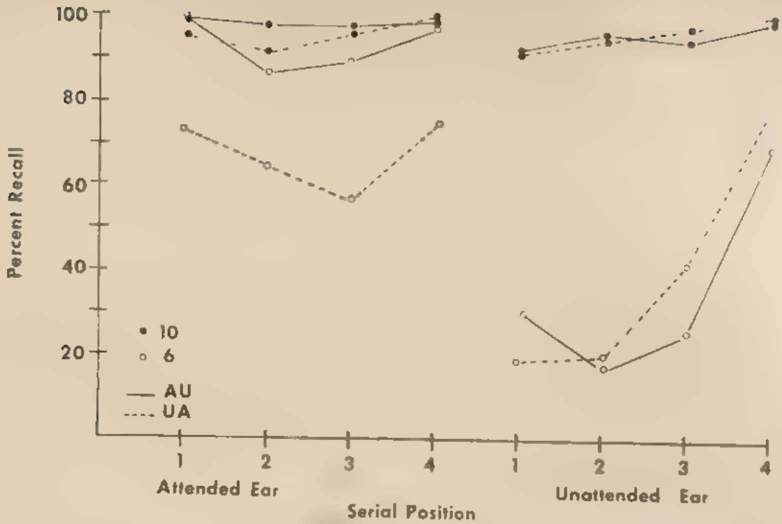


FIGURE 10. Percentage correct serial recall in Experiment IV as a function of digit span (10 or 6), attention (attended unattended ear), order of report (attended unattanded, AU; unattended attended, UA), and serial position

In the next experiment 4 *Ss* with low digit spans were forced to respond with 4 attended-ear and 4 unattended-ear digits in both AU and UA reports. This procedure permits a serial recall scoring analysis, the results of which can be compared with data from the high-span *Ss* in the last experiment.

Experiment IV

Method

Subjects and apparatus. Digit spans were assessed as in Experiment III and 4 *Ss* with digit spans of 6 were selected for participation as *Ss* in the present experiment. The apparatus was the same as that used in the previous experiments.

Procedure. Each *S* participated in 2 test sessions. In both sessions *Ss* received the test tape used in Experiment III, which consisted of 2 practice blocks (10 trials each) and 2 test blocks (30 trials each). Headphones were reversed in the second test session for all *Ss*.

In the first test session 2 *Ss* were instructed to attend to their right ear and 2 to their left ear. One *S* attending to each ear was given the report sequence AU–UA while the other *S* was given UA–AU. The instructions were the same as those previously administered, i.e., to report as many items as they could remember from both ears. As in the previous experiments *Ss* were instructed to report the items in the order in which they were presented. This test session was used to establish that the performance of these *Ss* was comparable to that of the low-span *Ss* in Experiment III.

In the second test session, *Ss* received the same test tape but with headphones reversed so that attended and unattended messages were different from those in the first session. Each *S* attended to the same ear (right or left) and received the same report sequence (AU–UA or UA–AU) as he did in the first session. The only change in the second session was in the report instructions. Each *S* was told that he was to report 4 attended-ear digits and 4 unattended-ear digits in both AU and UA reports. Subjects were told to guess if they did not remember the numbers. In addition they were instructed to make sure to report the numbers that they did remember in the correct serial position.

Results

The purpose of the first test session was simply to establish that the performance of *Ss* in the present experiment, when they were allowed to report in the manner of their choice, was comparable to that of the low-span *Ss* in Experiment III. The *Ss* in the present experiment responded with only 1 or 2 unattended-ear items in both AU and UA reports. Collapsed across *Ss* and order of report (AU and UA), their mean percentage correct recall on the unattended ear was 11.67%, 12.91%, 35%, and 83.33%, respectively, for Serial Posi-

tions 1–4. Comparison of these percentages with those shown in Figures 8 and 9 indicates that *S*s in the present experiment were comparable to the low-span *S*s in the previous experiment.

A serial recall scoring procedure was used for data in the second session, in which *S*s were required to report 4 attended-ear and 4 unattended-ear items on each trial. In this analysis a number was scored correct only if it was recalled in the correct serial position. Given an attended-ear message, 5, 9, 1, 3, and an unattended message, 4, 2, 6, 8, with an AU report instruction, if *S* recalled 5, 1, 9, 3, 4, 6, 2, 8, he was given credit only for the first and fourth serial position digits in both ears. In the free-recall scoring procedure used previously, *S*s would have received perfect credit for this report.

Raw scores from the second session were converted to percentage correct recall, and they are plotted in Figure 10. Also shown in Figure 10 is a serial recall analysis of the data for the 4 *S*s with a 10-digit span who participated in Experiment III. As these 4 *S*s all responded with 4 attended-ear and 4 unattended-ear digits on each trial, their data were simply rescored. Comparison of the data in Figure 10 with those shown in Figures 8 and 9 reveals 2 interesting findings. First, the performance of high-span *S*s remained unchanged with a serial recall scoring analysis, indicating that these *S*s were not simply guessing more items than low-span *S*s. Second, low-span *S*s were able to recall only 1 or 2 unattended ear digits in the correct position, even though they responded with 4 digits on each trial. Together these findings suggest that it is digit-span ability, and not a response strategy correlated with digit span, which is the important variable underlying differential performance.

Discussion

Several group findings were revealed reliably by analyses of variance across Experiments I, II, and III: (*a*) attended-ear recall was superior to unattended-ear recall; (*b*) attended-ear serial position functions were relatively flat, while positively increasing serial position curves characterized unattended-ear performance; (*c*) attended-ear recall was impaired when reported second relative to when reported first; (*d*) unattended-ear recall accuracy did not vary as a function of order of report; and (*e*) AU reports were superior to UA reports.

TABLE 4

Mean Number of Unattended Items Recalled Correctly per List (Collapsed over Order of Report) as a Function of Digit Span

Digit span	*M* no. of unattended digits recalled
5	1.14
6	1.78
7	1.78
8	2.80
9	3.46
10	3.80

However, of these experimental findings, only 3 were found to be reliable across *S*s of all digit-span groups: (*a*) impairment of attended-ear recall when reported second, (*b*) invariance of unattended-ear recall accuracy as a function of report order, and (*c*) superiority of AU reports. There was a pronounced absence of Serial Position × Attention interactions in high-span groups, and the tendency for attended-ear superiority was diminished.

What are the mechanisms or processes underlying digit span which result in more accurate unattended-ear performance when digit span is high (8, 9, and 10) than when it is low (5, 6, and 7)? There are at least 2 alternatives. First, we have defined active storage in terms of attention and rehearsal processes and echoic storage as an auditory buffer in which information is relatively unanalyzed. We have also identified active storage with attended ear and echoic storage with unattended ear. In terms of this model, the more accurate unattended-ear performance of *S*s with high digit span reflects echoic stores of larger capacity. The mean number of unattended items reported correctly per list (collapsed over report order) was calculated for *S*s of all digit-span groups. These figures, shown in Table 4, indicate values of echoic storage capacity ranging from 1.14 digits (Span 5) to 3.80 digits (Span 10). Bryden (1971) hypothesized that echoic storage was limited to 1–2 items. While this

estimate adequately describes performance for *S*s with digit spans of 5, 6, and 7, it is too low for *S*s with spans of 8, 9, and 10.

The identification of active storage with attended ear and echoic storage with unattended ear implies that no attention is directed towards the unattended ear. If, as a second alternative, we modify our working model to include the possibility that some attention is directed towards the unattended ear, a different interpretation of the more accurate unattended-ear performance of high-span *S*s arises. Consider models in which "attention" is used synonymously with "central processing capacity," "energy," and/or "mental effort" (e.g., Kahneman, 1973; Moray, 1967; Posner & Boies, 1971). In models of this type a finite central processing capacity or amount of energy is available for information processing. Any mental operations such as "encoding," "retrieval," "rehearsal," and the like, which are thought to require central processing capacity, must share from a common source. Consequently, if at any particular moment there is a large investment of central processing capacity to one mental operation (e.g., retrieval), the amount of capacity remaining for other operations (e.g., encoding and rehearsal) is reduced. If the capacity available for additional mental operations is not sufficient, impairment of performance will result.

Consider the implications which this type of conceptual analysis has for dichotic memory. All *S*s in the present experiments were instructed to "attend" (invest central processing capacity) to a 4-digit message presented to one ear. If we assume that the amount of central processing capacity required by this task is a function of the relation between length of message and digit span, *S*s with a low digit span (5 or 6) would, in order to maintain accurate performance on the task, be required to invest more processing capacity than would *S*s with a high digit span (9 or 10). More spare capacity would thus be available to high-span *S*s. Given instructions to attend to one ear, but to recall messages from both ears, an *S* might be expected to direct his spare capacity to his unattended ear. Thus in terms of the original model, unattended-ear performance would be mediated by both active and echoic storage systems. Viewed in this way, the difference in unattended-ear performance detected between *S*s with and low digit spans reflects not a diffe in echoic storage capacities, but rathe greater contribution of active storage with high digit span.

It is not possible to decide at present which of these theoretical alternatives is mo sible. While the knowledge that 79 of the variability in dichotic memo accounted for by digit span does no the theoretical issues, it does increas mental precision and it points out d for future research.

REFERENCES

Broadbent, D. E. The role of auditory loc tion in attention and memory span. *Journal of Experimental Psychology*, 1954, **47**, 191–196.

Broadbent, D. E. Immediate memory and simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 1957, **9**, 1–11.

Broadbent, D. E. *Decision and stress.* London: Academic Press, 1971.

Bryden, M. P. Order of report in dichotic listening. *Canadian Journal of Psychology*, 1962, **16**, 291–299.

Bryden, M. P. The manipulation of strategies of report in dichotic listening. *Canadian Journal of Psychology*, 1964, **18**, 126–138.

Bryden, M. P. Attentional strategies and short-term memory in dichotic listening. (*Psychology*, 1971, **2**, 99–116.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Kahneman, D. *Attention and effort.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Kroll, N. E. A., Parkinson, S. R., & Parks, T. E. Sensory and active storage of compound visual and auditory stimuli. *Journal of Experimental Psychology*, 1972, **95**, 32–38.

Miller, G. A. The magical number 7 ± 2. *Psychological Review*, 1956, **63**, 81–89.

Moray, N. Where is capacity limited? A survey and a model. *Acta Psychologica*, 1967, **27**, 84–92.

Murray, D. J., & Hitchcock, C. H. Attention and storage in dichotic listening. *Journal of Experimental Psychology*, 1969, **81**, 164–169.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Posner, M. I., & Boies, S. J. Components of attention. *Psychological Review*, 1971, **78**, 391–408.

Sternberg, S. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

(Received March 26, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 81–87

AROUSAL AND THE RANGE OF CUE UTILIZATION [1]

STEPHEN J. BACON [2]

Indiana University

The present experiment explores the effects of arousal on the range of stimulus responsiveness in terms of sensitivity and response criterion changes, the amount of attention initially focused on the stimulus, and the locus of effect of arousal. The *S*s simultaneously performed a pursuit-rotor tracking task and an auditory signal detection task under different levels of arousal, temporal locus of effect conditions, and degrees of attention being focused on the auditory cue. Results indicate that arousal narrows the range of cues processed by systematically reducing responsiveness to those aspects of the situation which initially attract a lesser degree of attentional focus. This stimulus loss under arousal represents, independently of any response criterion changes, an actual diminution in the *S*s' sensitivity. In addition, it seems that arousal mediates its effect not so much by impeding the initial sensory impression as by affecting the capacity limitations and attentional control processes operating within short-term memory.

Discussions of the effects of emotional arousal on the breadth of attention are assuming an increasingly prominent role in accounts of the experiential and behavioral consequences of arousal (Easterbrook, 1959; Korchin, 1964; Wachtel, 1967; Wine, 1971). The basic theoretical formulation is provided by Easterbrook, who suggests that the effect of emotional arousal on attention is to narrow and focus the attentional field by systematically reducing the range of cue utilization. More specifically, it is proposed that responsivity to peripheral or less relevant stimuli is diminished, while responsivity to central, immediately relevant or dominant cues is maintained, if not in fact augmented. Empirical support for the Easterbrook hypothesis may be found in the experiments of Bahrick, Fitts, and Rankin (1952), Bursill (1958), and Wachtel (1968), which have shown that, while simultaneously performing a central pursuit-meter tracking task and a peripheral visual detection task, aroused *S*s evidence considerable decrement in responding to the visual stimuli.

Perhaps the primary question that can be put to the arousal–reduced cue-utilization formulation concerns the nature of the stimulus loss. The implicit assumption being made is that this involves a reduction in the range of actual sensitivity to stimuli. The possibility exists, however, that emotionally aroused *S*s do not really experience attenuation of sensitivity to peripheral or less relevant cues, but rather are more cautious in reporting the occurrence of these stimuli. The present experiment employs a signal detection methodology (Green & Swets, 1966) to determine whether reduced responsivity under arousal to peripheral stimuli reflects an actual attenuation of sensitivity to the signal, or merely an increase in the subjective decision criterion, so that it becomes less likely that the signal will meet the stricter criterion.

It should be noted that the main thrust of the literature has been directed towards investigating the effects of arousal in reducing responsivity to peripheral cues. Easterbrook (1959), however, had postulated that the narrowing and focusing of the attentional field under arousal involves not only diminished utilization of periph-

[1] This article is based on a dissertation submitted to the University of California, Los Angeles, in partial satisfaction of the requirements for the PhD degree. The author wishes to express his appreciation to his dissertation committee chairman, William E. Broen, and to the committee members, Eliot H. Rodnick, Lowell H. Storms, James P. Thomas, and Daniel M. Wilner. The author also wishes to thank Richard H. Price for his helpful comments.

[2] Requests for reprints should be sent to Stephen J. Bacon, Department of Psychology, Indiana University, Psychology Building, Bloomington, Indiana 47401.

eral cues, but also the maintained, if not improved, use of central, immediately relevant cues. This suggests 2 related hypotheses to be tested: (*a*) There is a differential shift in responsivity under arousal depending upon the initial amount of attention being focused on a cue; i.e., cues which initially attract less attention should show further diminished attention under arousal, while cues which occupy the primary focus of attention should perhaps show enhanced attention under arousal. (*b*) There is a heightened sensitivity under arousal to those cues on which a high degree of attention was initially focused.

In general, the evidence that arousal increases the performance of a central task is highly inconsistent. Moreover, the type of task that has been used to measure the proposition that arousal might augment awareness of central, primary stimuli (e.g., pursuit meter) typically involves complex perceptual–motor coordination and higher order cognitive organization, and is far from testing simple stimulus responsivity. Another purpose, then, of the present investigation is to study, under arousal, differential shifts in stimulus responsivity to task cues which diverge along the dimension of amount of attention (high vs. low) being focused on them.

Finally, the research reported here is also directed towards the question of the locus of the effect of arousal on the range of stimuli that are processed. The Easterbrook (1959) hypothesis essentially asserts that the effect of arousal involves capacity limitations in information processing and the activation of attentional control mechanisms. One of the fundamental issues in information processing concerns whether such capacity limitations and attentional effects operate during the initial perceptual encoding stages of processing, or whether they operate exclusively in short-term memory store following perceptual processing (Shiffrin & Geisler, 1973). The present study, by manipulating whether the response to target stimuli is given immediately or is given after a delay, explores whether the effects of arousal on stimulus responsivity occur at the locus of perceptual or memory processes.

Method

Subjects

The *S*s were 20 students at the University of California, Los Angeles, who were paid for their participation in the experiment. They ranged in age from 19 to 25 yr. and had an average age of 21.6 yr. (*SD* = 1.82).

Apparatus and Instrumentation

The experimental task consisted of the simultaneous performance of pursuit-rotor (PR) tracking and auditory signal detection. The fluorescent light target of a Lafayette pursuit rotor, set at 40 rpm, was tracked by means of a stylus containing a photoelectric cell. A Lafayette timer clock measured time on target (TOT) to the nearest .01 sec.

The auditory signal consisted of a 1,000-cps tone generated by a Hewlett-Packard audio oscillator and monitored through a Hewlett-Packard variable attenuator. The signal was activated by a 60-msec. pulse former triggered by a Davis timer. This timer was in turn programmed by a trial stepping device which controlled whether or not the signal was to be presented. On trials in which the signal was presented, the tone was then fed through a 40-db. fixed attenuator into a Grason-Stadler noise generator and mixed there with white noise. The output from the mixer, consisting then of either a signal embedded in the noise background or just the noise background, was then fed through a 30-db. fixed attenuator into a relay system and then into a set of Sharpe stereo headphones. The purpose of the 2 fixed attenuators was to shut out switching clicks from the mechanical program equipment. The programming equipment consisted of a series of interconnected Davis timers which automatically set trial intervals (7 sec.), time of signal presentation, and intertrial intervals (5 sec.). A switch selected whether the trial timer pulsed the noise or the PR. It should be noted that when the timer triggered the relay to allow the noise (or signal plus noise) to be on for a 7-sec. trial, the PR ran continuously. In this case, the TOT counter was connected to the relay so that it marked PR tracking time only when the noise was on (i.e., only during the 7-sec. trial interval). On the other hand, in those conditions in which the trial timer pulsed the PR for 7 sec., the noise was continually fed through to the headphones. In this case, TOT could still only be scored for 7 sec., since with the PR light turned off, the photoelectric cell in the stylus was not activated. The stimulus delay timer was controlled by a device which selected the particular delay interval for the stimulus presentation. In this experiment, delays of .5, 1.0, 2.0, 6.0, and 6.5 sec. were possible.

A public address system permitted 2-way communication between the *S*'s room and the room

containing the program equipment in which E was located. A Beede electrical stimulator generated electric shocks, which were delivered via penny electrodes taped to the fingers of the S's nonpreferred hand. After each experimental condition, the Ss filled out a 20-item, modified version of Spielberger's State Anxiety Inventory (SAI; Spielberger, Gorsuch, & Lushene, 1970).

Procedure

Determination of signal intensity. At the beginning of the session, the auditory threshold for each S for a 1,000-cps, 60-msec. tone, embedded after a .5-sec. delay in a 1.0-sec. noise burst, was determined by the method of limits. Ten alternating descending and ascending runs were given. Only the last 6 runs were used in the final calculation of the threshold. The signal level was then increased 3 db. above the established threshold.

Signal detection and PR tasks. In each condition, each S was given 100 trials lasting 7 sec. each, with an intertrial interval of 5 sec. During each 7-sec. trial, the S was instructed to perform the PR tracking task while simultaneously listening through a set of headphones for the occurrence of an auditory signal embedded within a white-noise background. The S was told that on each trial the signal might or might not occur, but that if it did occur, it could do so anywhere within the trial. He was informed that this was all totally randomized, was advised of the "gambler's fallacy," and was told to make an independent judgment on each trial. When the trial was over, the S rated his subjective confidence that a signal did or did not occur on a 4-point rating scale (signal positively there, probably there, probably not there, positively not there). The S was also asked to rate from 0% to 100% how hard he was concentrating on doing the PR tracking. The program for the signal presentation was as follows: randomly, but equally, distributed over 2 blocks of 50 trials each were 13 signals presented after a delay of 1 sec. into the trial, 12 signals after 2 sec., 12 signals after 6 sec., and 13 signals after 6.5 sec. Randomly interspersed within this program were 50 noise-only trials. Half of these were randomly designated as the noise trials to be used with the early signal presentation, and half were randomly designated as the noise trials to be used with the late signal presentations. At the beginning of each session, each S received 62 practice trials. Signal previews at each of the temporal delays were presented well above threshold before each block of trials. Rest periods of 5 min. each were given after the practice trials and after each block of 50 experimental trials, and a 15-min. rest period separated the nonarousal and arousal conditions for each S.

Level of arousal conditions. Each S performed the PR – signal detection task under counterbalanced no-shock and shock conditions. The nonarousal condition consisted of a no-shock PR – signal detection task. In the arousal condition, the S was informed that E was interested in studying the effects of stress, and that shock applied to the fingers of the nonpreferred hand would be used to induce stress. The level of shock that was perceived by the S as the most painful he was willing to endure was determined. The S was told that shocks would be randomly delivered while he was performing the task, so that the longer the interval between successive shocks, the more intense the shock would be. In each block of 50 trials, 20 shocks were delivered according to a fixed random schedule. Shocks were thus delivered during both intertrial interval and experimental trials, but never at the time of signal occurrence. Actual delivered intensities varied around the predetermined level as a function of intershock interval. The Ss were asked to complete the 20-item SAI after both no-shock and shock conditions.

Focus-of-attention conditions. The Ss were randomly assigned to one of 2 groups which differed in the amount of attention that was being focused on the auditory task. Degree of focus of attention on the auditory cue was manipulated by physical and instructional means designed to pull attention either towards or away from the auditory signals. When the auditory cue was to occupy the primary focus of attention, the PR was run continuously, and the noise coming over the headphones was pulsed for 7 sec. The S was told that when the noise goes on he was to begin PR tracking and listening for the auditory signal, and as soon as the burst ended he was to stop tracking and make his signal detection and concentration ratings. Thus, the auditory channel gained further importance in that it set the stage for task performance. The S was also told that, while he was required to perform both tasks simultaneously, the more important task was the auditory signal detection task. Furthermore, if he performed above a certain criterion level on the auditory task, he could receive a bonus of $1.00. The only qualification to this was that he had to perform the PR task to a criterion, although he was assured that the criterion was fairly low. All Ss were, in fact, rewarded. Following the practice trials, S was informed that his PR performance was adequate, and he was encouraged to maintain this level. After experimental trials 25, 50, and 75, S was again assured that his PR performance was adequate, was again reminded of the importance of the auditory task, and was encouraged to concentrate on listening for the signal. When the auditory cue was to attract less attention, the manipulations favored the PR task. Thus, the auditory noise background was on continuously, and the 7-sec. trial interval was determined by the pulsing of the PR, i.e., the target was illuminated and rotated for 7 sec. and then stopped and turned off. Pursuit-rotor tracking was emphasized, and a $1.00 bonus could be received for performing PR tracking above a criterion level, provided that minimal criterion performance be maintained on the auditory detection task. Following the practice

TABLE 1

Mean Sensitivity and Mean Slope Scores for No-Shock and Shock Conditions over High and Low Degree of Focus of Attention on the Auditory Cue (HF, LF), and Immediate and Delayed Response (RI, RD) Conditions

Conditions	Sensitivity scores		Slope scores	
	No shock	Shock	No shock	Shock
HF–RI	3.02	2.63	.53	.80
HF–RD	3.32	2.61	.74	.61
LF–RI	3.13	2.30	.65	.68
LF–RD	3.61	1.60	.42	.98

trials and at the given points during the experimental sessions, S was advised that his auditory detection performance was adequate, was encouraged to maintain this level, and was reminded of the primary importance of the PR task.

Temporal locus of effect condition. Temporal factors were manipulated by varying the time between auditory signal presentation and S's report, so that the response was either immediate or delayed. The trials containing the signals presented after 6- and 6.5-sec. delay, along with their noise-only trials, were collapsed together to make up the response-immediate condition. The signal presentations occurring 1 and 2 sec. into the trial, requiring S to delay his response 5 or 6 sec., taken together with their noise-only trials, comprised the response-delayed condition.

Results

The interpretation of the effects of arousal on the range of cue utilization depends upon the effectiveness of the no-shock vs. shock manipulation in producing differential levels of arousal. Raw scores for each S on the SAI were tabulated and then translated into the appropriate percentile rank scores given from the norms on college undergraduates (Spielberger et al., 1970). An analysis of variance performed on the arousal level scores revealed the mean difference between no-shock and shock conditions (Percentile Rank 58 vs. 91) to be hightly significant, F (1, 16) = 37.08, $p < .001$. No other differences between treatment conditions were found.

The actual sensitivity of the various treatment groups to the auditory signal is expressed in terms of a linear function, Δm, representing the measure of underlying sensory capabilities, independent of response bias. Green and Swets (1966) have proposed this index of sensitivity to be used instead of d' when the assumption of equal variance in the signal and noise distributions may not be met. Values of Δm were obtained for each S based on the rating data for each of the conditions in which S participated. Cumulative conditional probabilities of hits and false alarms at each of the decision criterion points were calculated, transformed to their corresponding Z score equivalents, plotted against each other, and, by the method of least squares, they were fit with a linear function. The Z score values of the X intercepts of the linear functions served as the estimates of Δm (Green & Swets, 1966). The analysis of variance conducted on these data showed that arousal clearly diminishes sensitivity to the auditory cue, F (1, 16) = 11.32, $p < .005$. Table 1 shows the mean sensitivity scores for the focus-of-attention, temporal factors, and arousal-level treatment combinations. It can be seen that arousal produces greater sensitivity loss when the auditory cue attracts less attention, and particularly when the response to it is delayed. The planned comparisons performed to test whether arousal increases sensitivity to the auditory cue when it occupies the primary focus of attention and response is immediate or delayed, and whether arousal reduces sensitivity to the auditory cue in the low-attention condition for immediate and delayed responding, revealed that the primary effect of shock was to reduce sensitivity in the low-attention – response-delayed condition, F (1, 16) = 9.61, $p < .01$. The other comparisons fell well short of significance: F (1, 16) < 1 for the primary-attention – response-immediate condition; F (1, 16) = 1.17, $p >$ .25, for the primary-attention–response-delayed condition; and F (1, 16) = 1.65, $.25 > p > .10$, for the low-attention – response-immediate condition.

Another independent sensory parameter is the slope value of the linear function. The traditional interpretation of the slope in signal detection theory is that it equals the ratio of the variance of the noise dis-

tribution to that of the signal distribution. An analysis of variance was carried out on the slope values. A significant main effect was found for arousal level, F (1, 16) = 4.91, $p < .05$, the mean value of the slope increasing from .58 in no-shock conditions to .77 with shock. A significant interaction for arousal level and order effects, F (1, 16) = 6.60, $p < .025$, suggests that the main effect of increased slope for aroused *S*s was due primarily to those *S*s in the no-shock-first, shock-second order condition. A significant 3-way interaction of focus of attention, temporal factors, and arousal level was also obtained, F (1, 16) = 5.29, $p < .05$, and the mean slope values for each of these treatment combinations are also presented in Table 1. The planned comparisons indicated that shock significantly increases the slope value only in the low-attention – response-delayed condition, F (1, 16) = 7.66, $p < .025$.

Typically, changes in the ratio of the variances of the noise to the signal distribution have been taken to indicate changes in the way *S* is processing information about signal parameters. However, while it is of interest to know in the present experiment that not only sensitivity changes but also underlying information-processing strategy changes occur under arousal, no additionally meaningful psychological information can be ascertained from the slope statistic concerning the locus of arousal's effect. There will, therefore, be no further discussion of the slope data.

Another measure which further describes *S*'s behavior on the auditory detection task is the hit rate *Z* score, $Z(y|SN)$, where y = yes, signal was present, and SN = signal was added to noise background, which provides fairly direct estimates of shifts in response criterion when sensitivity values do not differ. In addition, it is the hit rate scores in this study which are most closely analogous to the perceptual performance measures of previous central-tracking – peripheral-light studies. Table 2 shows the mean hit rate *Z* scores at each of the cumulative subjective criterion

TABLE 2

Mean Hit Rate *Z* Scores at Each Criterion Point for No-Shock and Shock Conditions over High and Low Degree of Focus of Attention on the Auditory Cue (HF, LF) and Immediate and Delayed Response (RI, RD) Conditions

Conditions	No shock			Shock		
	Strict	Moderate	Lax	Strict	Moderate	Lax
HF–RI	−.531	.512	1.296	−.260	.592	1.320
HF–RD	−.348	.690	1.661	−.179	.756	1.487
LF–RI	−.305	.573	1.252	−.491	.371	.848
LF–RD	−.229	.546	.925	−.689	.629	.962

points as a function of focus of attention, temporal factors, and arousal level. The analysis of variance and the planned comparisons performed on the overall mean hit rate *Z* score, $\bar{Z}(y|SN)$ revealed no significant differences between treatment conditions. This finding is somewhat surprising in that one might expect that, given the observed reduction in sensitivity under shock, an overall lowering of the average number of hits should follow. An explanation of this finding, which seems to be supported by reference to Table 2, is that treatment manipulations had opposing effects at different criterion points, thus leaving the overall average number of hits unchanged.

Particularly noteworthy in Table 2 are the hit rate *Z* score data associated with the most stringent criterion point, in that the previous central-tracking – peripheral-light experiments, which required *S* to switch off the peripheral light when he saw it was on and recorded the amount of time the light remained on, were probably dealing with a criterion range covered by the most stringent criterion point in the present investigation. Table 2 shows that, at the most stringent criterion level, under arousal there is a moderate increase in hit rate when the auditory cue occupies the primary focus of attention, but a marked decrease in the hit rate when the auditory cue attracts less attention, most dramatically when response is delayed.

Time on target per trial performance on the PR was also evaluated by an analysis

of variance. The results that merit consideration here are the significant Temporal Factor × Arousal Level interaction, $F(1, 16) = 4.99$, $p < .05$, which indicates that only when the response to the auditory cue is delayed did shock produce decrement in PR performance. The planned comparisons revealed that shock produced significantly lower PR scores only in the low attention to the auditory cue (hence centered attention on the PR)–response-delayed condition, $F(1, 16) = 8.42$, $p < .025$.

Discussion

The results of the present experiment support Easterbrook's (1959) hypothesis that the effect of arousal is to reduce the range of cue utilization, and the results are also consistent with the conclusions of Bahrick et al. (1952), Bursill (1958), and Wachtel (1968) that emotional arousal decreases responsiveness to peripheral stimuli. More important, the data indicate that the stimulus loss under arousal involves a diminution in *S*'s sensitivity, independently of any response criterion changes. Moreover, arousal seems to have differential effects depending upon the degree of attention the stimuli attract, sensitivity loss systematically occurring to those cues which initially attract less attention. On the other hand, the present study offers no support for the suggestion that arousal might augment sensitivity to cues which occupy the primary focus of attention. One must note, however, that the PR is such a demanding and compelling task that, despite the manipulations designed to focus attention on the auditory task, it is quite likely that the auditory cue never achieved attentional centrality. Thus the present study may not have provided the optimal conditions for a central enhancement effect to emerge.

Some additional evidence emerges from this study which may be viewed as an extension of the findings that arousal reduces the range of cue utilization. A significant decrement in both PR performance and sensitivity to the auditory cue was obtained under arousal in the low-attention – response-delayed condition. This seems to suggest that when *S* is required to perform the PR task while also retaining the memory trace of the auditory signal, arousal reduces *S*'s ability to attend to both tasks simultaneously. This provides support for the notion that arousal reduces the number of cues *S* can handle at [a] time, the capacity, in Korchin's (1964) terms, for dual activity. In addition, these results indicate that arousal introduces further capacity limitations within the short-term memory system.

The response-immediate and response-delayed manipulations in the present study were directed at determining the locus of effect of arousal in reducing the range of stimuli that are processed. The response-immediate condition was designed to test the effects of arousal on perception. It is realized that response delays of .5 and 1.0 sec. were involved here and thus memory mechanisms may be implicated. Nonetheless, if arousal affects perception of external stimuli by attenuating signals arriving over sensory channels, this would surely have resulted in a sensitivity loss in the response-immediate condition. While a trend towards diminished sensitivity was found in the low-attention – response-immediate – shock condition, it did not attain statistical significance. Thus, there is little evidence from the present investigation that arousal acts effectively to impede perception of stimulus input. On the other hand, the response-delayed condition, designed to evaluate the effects of arousal on memory processes, did show reduced sensitivity to the auditory cue under low-attention – shock conditions. This suggests that arousal narrows the range of stimuli that are processed by impairing the memory traces of those signals which initially attract less attention. Arousal may exert its effect in this condition by both overloading the system and thereby exceeding the capacity limitations of short-term storage, and by interfering with the attentional control processes, such as rehearsal, that aid in maintaining information in short-term memory. Within this context, it is interesting to note that the differential effects of the focus-of-attention manipulations in the present experiment become evident only within memory and under the capacity limitations imposed by arousal. The results and conclusions presented here are consistent with those advanced by Shiffrin and Geisler (1973), that attentional effects and capacity limitations operate solely within short-term memory storage following perceptual processing.

REFERENCES

BAHRICK, H. P., FITTS, P. M., & RANKIN, R. E. Effects of incentives upon reactions to peripheral stimuli. *Journal of Experimental Psychology*, 1952, 44, 404–406.

BURSILL, A. E. The restriction of peripheral vision during exposure to hot and humid conditions. *Quarterly Journal of Experimental Psychology*, 1958, 10, 113–139.

EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 1959, 66, 183–201.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

KORCHIN, S. J. Anxiety and cognition. In C. Scheerer (Ed.), *Cognition: Theory, research, promise*. New York: Harper & Row, 1964.

SHIFFRIN, R. M., & GEISLER, W. S. Visual recognition in a theory of information processing. In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology: The Loyola symposium*. Washington, D.C.: Winston, 1973.

SPIELBERGER, C. D., GORSUCH, R. L., & LUSHENE, R. E. *The state-trait anxiety inventory: Test manual for form X*. Palo Alto, Calif.: Consulting Psychologists Press, 1970.

WACHTEL, P. L. Conceptions of broad and narrow attention. *Psychological Bulletin*, 1967, 68, 417–429.

WACHTEL, P. L. Anxiety, attention, and coping with threat. *Journal of Abnormal Psychology*, 1968, 73, 137–143.

WINE, J. Test anxiety and direction of attention. *Psychological Bulletin*, 1971, 76, 92–104.

(Received March 27, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 88–94

ARTICULATORY INTERFERENCE AND THE MOW... HETEROPHONE EFFECT[1]

JOHN L. BRADSHAW[2] AND NORMAN C. NETTLETON

Monash University, Clayton, Victoria, Australia

When *S*s are required to utter rapidly a sequence of visually presented which are closely similar in spelling, but not in pronunciation, such a DOWN or HOME–SOME, hesitations and confusions may result. These were found to be less apparent when the same items are so interleav members of "heterophone" pairs such as the above are no longer adja each other. It was shown that this heterophone effect was unlikely to d on the *S*s scanning ahead or perceiving later words in peripheral vision phenomenon was studied further in terms of the times for initiating and pleting the articulation of items either separately or in pairs. By varyi contingencies of presentation and utterance of paired items (hetero and neutrals), it was found that a pronunciation set could be establis cause subsequent interference. However, overt articulation was nec for this, not merely covert reading or memorizing, and the effect was fested in the time taken to initiate articulation, not its duration. The fin were discussed within the context of word recognition, pronunciation set grapheme–phoneme recoding.

A phenomenon in English orthography is the existence of large groups of words such as MOWN–DOWN, QUART–PART, HORSE–WORSE, HOME–SOME, and TOW–COW. Such heterophone pairs share largely the same spelling, but with some considerable pronunciation differences. When we encounter any one such word in isolation, there is usually little difficulty in pronunciation. When a number are juxtaposed, the speaker may become confused.

A number of experiments are reported where this phenomenon is investigated. The findings are later considered within the context of phonemic recoding of visually presented stimuli ("indirect access" theories, Smith, 1971), Morton's logogen theory (Morton, 1969), sensory and motor "buffer" stores in reading, and the role of peripheral vision in reading.

Experiment I

The aim of this experiment was to examine the possibility that *S*s may have more difficulty in correctly pronouncing a sequence of words where the members of heterophone pairs such as ROOT–FOOT, RAID–SAID, BOWL–FOWL, etc., are adjacent to each other, than where they are separated from each other by single members of other heterophone pairs.

Method

Material. Dictionaries were searched to obtain 48 pairs of heterophone items. The words varied from 3 to 8 letters in length, and all except 9 pairs were single syllables, the remainder being bisyllabic. The items were arranged in 4 vertically organized lists. In the first list the 2 members of a pair were always one immediately above the other, i.e., there were no intervening items (List 0). In the next, the same words were arranged in an interleaved form, so that there was always one item between the 2 members of a heterophone pair—these members themselves serving to separate other heterophone pairs—thus: HOME–NORTH–SOME–WORTH–SOUGHT–WEAR–DROUGHT–FEAR, etc. There was therefore one intervening item (List 1). In the third list, the same words were arranged so that there were 3 intervening items (List 3), thus: HOME–NORTH–SOUGHT–WEAR–SOME–WORTH–DROUGHT–FEAR, etc. Finally, in the last list (List 7), the words were so interleaved that there were always 7 intervening items, e.g., HOME–NORTH–SOUGHT–WEAR–ROLL–WEIGHT–ROOT–THOUGH–SOME–WORTH–DROUGHT, etc.

Procedure. These 4 lists can be arranged in 24 possible orders. Twelve of these were selected which permitted 3 different complementary designs of the Latin square type. The 12 *S*s were all required to read aloud each of the 4 lists twice, i.e., in accordance

[1] This research was supported by an award from the Australian Research Grants Committee. We wish to acknowledge the assistance of Kevin Donahoo and Robert Wood in the design and construction of apparatus.

[2] Requests for reprints should be sent to John L. Bradshaw, Department of Psychology, Monash University, Clayton, Victoria 3168, Australia.

the particular specified sequence and also in the ...se of that sequence, thus: 3, 0, 1, 7, 7, 1, 0, 3. Th... were timed with a stopwatch while reading, a... were encouraged to read as fast as possible, ke...ing mistakes of pronunciation and hesitations to... minimum. They were instructed to continue re...ng if they did make a mistake, and not to go b... and correct it. The *S*s, 4 males and 8 females, we... staff or postgraduate student members of the de...artment of psychology of Monash University.

...esults and Discussion

The means of the reading times for the 4 ...sts, across the 12 *S*s, were: List 0, 41.5 ...ec.; List 1, 38.3 sec.; List 3, 36.7 sec.; ...nd List 7, 37.0 sec. A 2-way analysis of variance (within *S*s) showed a significant difference between the lists, $F\ (3, 33) = 34.1$, $p < .001$. In fact List 0 was the longest for all 12 *S*s, without exception, and with only one exception, List 1 took a longer time to read than List 3.

It seems therefore that when *S*s are ...uired to read lists of such heterophone items, as fast as possible consistent with a minimum of mistakes, there is considerable interference both when the items are adjacent, and even, although to a lesser extent, when they are interleaved so that they are each separated by a "neutral" item.

A number of possible explanations will be discussed later. One possibility, otherwise of little theoretical interest, is simply that *S*s were either scanning ahead or, what comes to the same thing, able to see the next member of the pair in peripheral vision. This might have caused interference in the case of adjacent heterophones.

EXPERIMENT II

There is perhaps a fairly simple test to see whether a word beyond the one currently being processed can interfere with the current one. In English there are a number of heteronyms, such as LEAD (rhyming with either BEAD or DEAD, depending upon meaning), ROW, SOW, and BOW (rhyming with either HOW or LOW), TEAR (rhyming with either PEAR or FEAR), WIND (TINDER or BINDER), and WOUND (...OUNDED or BOUNDED). If *S* encounters ...EAD followed by BEAD (or DEAD) in a list, is he more likely to pronounce it in rhyming fashion? Lists were constructed, containing filler and critical items such as the above, to see whether scanning ahead (or peripheral vision) tends to influence the pronunciation of ambiguous heteronyms.

Method

Material. Two vertical lists of single words were constructed. Both contained identical filler items—rhyming pairs such as SINK–DRINK, STONE–BONE, etc., interspersed with heterophone pairs such as OUR–FOUR, NEVER LEVER, etc. In this way it was hoped to avoid the establishment of a set or expectancy for complete or partial rhyming. At certain positions 6 critical items were located. In one list these were BOW–COW, TEAR–FEAR, LEAD–BEAD, ROW–LOW, WIND–BINDER, WOUND–BOUNDED. In corresponding positions in the other list were found BOW–TOW, TEAR PEAR, LEAD–DEAD, ROW–HOW, WIND–TINDER, WOUND-WOUNDED.

Procedure. Each of the 9 *S*s, 3 males and 6 females, was required to read the 2 lists aloud, as though he was being timed. In fact it was noted whether or not the *S*s pronounced the critical, ambiguous heteronym item to rhyme with the following word. The *S*s were given the 2 lists one before and one after another experiment lasting approximately 1 hr. Half of the *S*s were given one of the 2 lists first, and the others were given the other list first.

Results

The critical items were scored as rhyming or not rhyming. If *S* pronounced either the first or second word of a pair incorrectly, this was scored as an error. The total number of rhymes, nonrhymes, and errors are shown for each list in Table 1, as are the totals over the 2 lists. Table 2 shows the number of times *S*s pronounced the given test word the same or a different way in each list. Table 3 shows how each test word was pronounced, if pronounced correctly, on a total of 18 presentations.

TABLE 1

TOTAL NUMBER OF RHYMES, NONRHYMES, AND ERRORS FOR EACH LIST, OVER NINE SUBJECTS

Pronunciation	First list	Second list	Total
Rhymed	29	22	51
Not rhymed	21	31	52
Errors	4	1	5

TABLE 2

Total Number of Times Subjects Pronounced a Word the Same Way in Both Lists

Test word	Same	Different	Errors
BOW	6	2	1
TEAR	8	1	0
LEAD	6	2	1
ROW	6	2	1
WIND	8	1	0
WOUND	6	1	2

Discussion

The results shown in Table 1 suggest that scanning ahead did not play an important part in the reading of these lists. Overall, *S*s were equally likely to rhyme or not rhyme the critical items. However, this may merely be due to a strong overriding tendency for *S*s to pronounce the ambiguous test word in one particular way. Table 2 shows that this was the case—the test words were pronounced the same way in both lists on 40 occasions and differently on 9. This does not mean that *S*s agreed on their preferred pronunciation. Table 3 shows that only 2 words—TEAR and WIND—were pronounced in the same way by a majority of the *S*s.

If scanning ahead does occur it does not appear to have sufficient influence to override *S*s' tendency to pronounce a given test word in their own preferred way. It therefore seems unlikely that the heterophone effect in Experiment I was due to *S*s' scanning ahead, or perceiving a later word in peripheral vision.

This experiment could be repeated with the ambiguous words (BOW, TEAR, LEAD, etc.) occurring after instead of before the words with which they could be rhymed (COW–BOW, TOW–BOW, etc.). This would show whether the first word encountered influences the second. However, there is evidence for this in the present data, because all 5 errors were due to *S*s' choosing a particular pronunciation of the first word and then making a nonword of the second word by rhyming it with the first, e.g. pronouncing the pair BOW and TOW as if to rhyme with COW. Thus the heterophone effect in Experiment I may be the influence of an earlier word on a later one, rather than the influence of one seen below in peripheral vision on the one currently being processed.

TABLE 3

Pronunciation and Errors for Each Test Word for the Two Lists Combined, over Nine Subjects

Test word	Pronounced as in:	No. of times	Errors
BOW	Tow	8	1
	Cow	9	
TEAR	Pear	15	0
	Fear	3	
LEAD	Dead	9	1
	Bead	8	
ROW	How	8	1
	Low	9	
WIND	Tinder	13	0
	Binder	5	
WOUND	Wounded	9	2
	Bounded	7	

Experiment III

In the first experiment, it was found that it took longer to read material correctly when the members of heterophone pairs were adjacent than when they were separated by intervening items. If only one member of a pair is articulated, the first being merely observed or subvocally identified, does the heterophone effect persist? If it does, it would imply that overt articulation may not be necessary to cause interference. Successive pairs of words were therefore presented on a computer-controlled oscilloscope, *S*s being required to observe the first and name the second, vocal initiation times being obtained.

Method

Material. Because of equipment and other constraints, words of only 4 letters were employed. A total of 25 heterophone pairs were obtained and a PDP-8/S computer was programmed to present them on a Tektronix RM 503 oscilloscope screen. Each item consisted of a pair of words, either a heterophone or a neutral pair. The latter consisted of one from each of 2 otherwise unrelated heterophone pairs. Thus the 2 heterophone pairs WART–PART and HIVE–GIVE generated the 2 neutral pairs WART–GIVE and HIVE–PART. Heterophone and neutral pairs were pseudorandomly intermingled, with certain constraints to avoid uneven sequencing effects. Every word appeared both at the beginning and end of a heterophone and a neutral pair, i.e., it

occurred somewhere in each of the 4 possible positions. Thus the 25 heterophone pairs generated 50 heterophone items, taking sequencing of members into account. There were also, correspondingly, 50 neutral items.

Procedure. The *S*s sat in front of the oscilloscope screen, at a distance of 40 cm., viewing it through a tachistoscope-type viewing hood. On each trial the *E* said "Ready" and *S* pressed a hand-held button. The first words then appeared in a faintly luminous target area for a total of 237 msec. In fact it was "written up" 3 times, each left-to-right scan taking 79 msec. After a blank interstimulus interval of 193 msec., the second word appeared in the same location, also for 237 msec. The *S*s were required to note the first word but to make no verbal response. They shouted out the second word as soon as they could, closing a voice-operated switch which stopped a timing routine begun at the offset of the second word. The initiation time thus obtained was printed out on the teletype. There were 8 new *S*s, 5 females and 3 males. Any stutters or articulatory blocks were excluded from analysis. However, over all *S*s, only 5 neutral and 7 heterophone items were so excluded.

Results and Discussion

The average time to initiate a verbal response (naming), after the offset of the second word (duration of 237 msec.) was 286 msec. for heterophone pairs and 288 msec. for neutrals. Thus there was a minimal and nonsignificant difference in a direction opposite to the one predicted, only 3 of the 8 *S*s producing longer vocal initiation times for heterophone pairs, $t\ (7) = .3$, $p > .8$. It would seem that under the conditions of this experiment, the second item of a heterophone pair can be initiated as readily as the second item of a neutral pair. Any slight facilitation for heterophone pairs may perhaps stem from easier decoding, resulting from the shared letters appearing spatially superimposed.

Experiment IV

When *S* covertly identified but did not articulate the first member of a pair of words, the expected heterophone effect failed to materialize in the initiation times for articulating the second word. Can it be made to reappear, under the conditions of separate paired presentations when *S* also articulates the first word and thus triggers the presentation of the second? In this experiment not only initiation times but also duration times were obtained for the articulation of the second word.

Method

Material and subjects. The same material and sequence were employed as in Experiment III. However, a totally new group of 8 volunteer *S*s, 4 males and 4 females who were entirely naive as to the material being used, was obtained.

Apparatus and procedure. The computer was programmed to present the first word of a pair for 237 msec. when *S* pressed the hand-held "ready" button. Exactly 388 msec. after *S* started to name the first word, the second word appeared, also for 237 msec. The computer and auxiliary equipment were programmed to time both initiation (after offset of the second word) and duration of *S*'s vocalization of the second word. The long interval (388 msec.) was chosen to insure that vocalization of the first word was completed before the second word went off and to insure that there was a gap between the 2 articulations. This permitted the timing of the parameters relating to the second word alone. Measurement of the duration of the second word was achieved by the construction of a special device. Since there are silent periods within articulated words, a voice-operated switch is not kept closed until the offset of the last voiced period, but tends to oscillate between its open and closed positions as it attempts to follow the intensity changes in the articulation. The voice-operated switch was therefore used to start a chronoscope and a specially constructed timer, which was reset each time it received a pulse from the voice-operated relay. If it did not receive a pulse this timer timed out after a preset interval (100 msec. in this experiment) and stopped the chronoscope, which thus gave the duration of the word plus 100 msec. The computer was programmed to reset the chronoscope and the timer at the commencement of each trial, and to hold them in this mode until the second word had been displayed. The initiation time was determined as before by a timing loop in the computer and printed out on the teletype. The *E* monitored *S* and the equipment and noted any trials on which a preliminary stutter was timed instead of the whole word. Over all *S*s, 13 neutral items and 14 heterophones were excluded from the analysis of initiation times, and 22 neutral items and 19 heterophones from the analysis of duration times. Some initial technical problems led to the exclusion of the extra duration times.

Results and Discussion

This time, all 8 *S*s showed faster initiation times to the second word with heterophone pairs than with neutral pairs, $t\ (7) = 4.9$, $p < .002$, again in the direction contrary to that predicted. The mean for heterophones was 273 msec. (after stimulus

offset), and for neutrals 289 msec. For some reason the trend observed in Experiment III became highly significant, although it should be noted that *S*s now had to articulate the first word. No difference in duration times was observed between heterophones (243 msec.) and neutrals (245 msec.), 4 *S*s showing very small differences in one direction, and 4 in the other. Obviously any effects operate upon the setting up of articulatory programs, not upon their operation.

The continued absence of the expected heterophone effect could possibly be due to the fact that the 2 words, although spatially superimposed, were temporally separated, as were the 2 verbal responses. It is important to see whether the strong effect obtained in serial list reading can be reproduced in the discrete situation of paired presentations.

Experiment V

In this experiment both words were presented in immediate succession, one above the other. The *S*s were required to articulate both together, and as before both initiation and duration times were measured. It was hoped that this way the predicted heterophone effect would reappear.

Method

Procedure. Using the same stimuli as before, and a new group of 8 naive *S*s, the 2 words were presented in immediate succession. The first word was written up twice and thus appeared for 158 msec. It was immediately followed by the second word, again for 158 msec., but this time directly below the first one, to avoid masking and facilitative effects which could result from superposition. The *S*s were required to name both words as fast as possible, both initiation and duration times being measured as before. In order to insure that the chronoscope measuring duration times was not stopped by a short pause between the 2 words, the timer was set to time out after 400 msec. The chronoscope thus gave the duration time for the 2 words plus 400 msec., a constant subsequently subtracted from the data.

Results and Discussion

In this experiment, with the initiation times the expected heterophone effect reappeared, the heterophones being on average 26 msec. longer than the neutrals, $t\ (7) = 2.02, p < .05$, one-tailed. Only one of the 8 *S*s gave a contrary result. The mean for heterophones was 276 msec., and for neutrals 250 msec. With the duration times, there was also an 11-msec. decrement with heterophones (701 msec.) as compared with neutrals (690 msec.), though this failed to reach significance, $t\ (7) = 1.7, p > .1$, being found only with 5 of the 8 *S*s. Thus the expected heterophone effect can be made to reappear with discrete presentations of paired stimuli. However, it is much weaker than that originally found with the continuous list-reading material. In Experiment I there was a difference of over 4.5 sec. between lists at the 2 extremes, each being composed of 48 pairs of heterophones and 48 neutrals. Thus there was in effect a difference of over 90 msec. per word pair. If the initiation and duration differences from Experiment V are added, there is a difference of little more than 38 msec. per pair. Consequently, it seems that much of the original heterophone effect stemmed from the continuous forced pace of articulation, not present with discrete presentations.

Experiment VI

Experiment V established some of the conditions necessary to demonstrate the heterophone effect with single word pairs. Whether actual articulation of both words is necessary for interference to occur is still uncertain, although this was the starting point for Experiment III. Experiment V was therefore repeated without articulation of the first word.

Method

Procedure. Exactly the same *S*s and procedure were used as in Experiment V except that instead of articulating both words, *S*s were required to note the first word, articulate the second as quickly as possible, and then, after the teletype had finished printing out the response time, tell *E* what the first word had been. There was an interval of approximately 3 sec. between presentation of the first word and *S*'s being free to report it. This depended on how long *S* took to respond with the second word. As before, both initiation and duration times were measured for the second word presented, i.e., the

first word called out by *S*. The timer controlling the chronoscope was set to 100 msec., and this value was subtracted from the chronoscope readings.

Results and Discussion

While the heterophones (508 msec.) still gave rise to longer initiation times than the neutrals (493 msec.), the difference (15 msec.) was smaller and 3 *S*s gave results in the opposite direction, making the effect nonsignificant. The duration times were nonsignificantly longer for the neutrals (349 msec.) than for the heterophones (342 msec.), $t = 1.7$, $p > .1$. The heterophone effect seems therefore to require *S* to articulate both members of the pair.

General Discussion

When people are required to read aloud a continuous list of material consisting of paired heterophones such as WART–PART, etc. (Experiment I), they find it more difficult than when the heterophone members are separated by "neutral" items (in reality, other heterophone members similarly separated). Some considerable residual difficulty persists even when there is one intervening item between successive members of a heterophone pair. The effect can also be found with discretely presented pairs of stimuli (Experiment V), when it is seen to depend largely upon the time taken to initiate the appropriate verbal response rather than the duration of that response. However, the cumulative, interactive effect of load with a continuous list appears important, the effect being very much weaker in the discrete mode of presentation (Experiment V), and nonexistent when the second word is presented only after *S* articulates the first (Experiment IV).

The effect demands the articulation of each item, and is not apparent in the initiation time for the second item when *S* merely notes the first item (Experiment III) or temporarily stores it for recall after pronouncing the second (Experiment VI). Moreover, the greater effect in the continuous list-reading task cannot be due solely to scanning ahead or to the employment of peripheral vision, since the pronunciation of ambiguous heteronyms such as LEAD, WIND, etc., is apparently not influenced by subsequent words such as BEAD–DEAD, TINDER–BINDER (Experiment II). Kolers and Lewis (1972) review the role of peripheral vision and eye movements in vision. They conclude that peripheral vision may indeed be more than simply a ranging instrument guiding the fovea. Some content analysis seems possible from input presented through several degrees of visual angle. However, this is successful only when all the information, foveal and peripheral, is part of a single word, and not 2 or more words distributed between the 2 areas. Kolers and Lewis conclude that "simultaneous processing of different linguistic units from center and periphery is not a strategy that the rapid reader could profitably engage in [p. 123]."

Failure to produce a heterophone effect in either of the 2 experiments where *S* merely "noted" or "stored" the first item (Experiments III and VI) seems to go against Morton's (1969) logogen theory. This states that any act of storage or identification of a word calls up a motor (articulatory) program, which would interfere with any other such item being processed. However, it is possible that in both these experiments the material persisted at a satisfactory and noninterfering level in some form of sensory "buffer" store.

It is debatable whether phonological recoding inevitably precedes or accompanies further processing—the indirect access theory (criticized by Smith, 1971). The latter probably owes its origins to work on short-term memory and acoustic confusions. (Baddeley, 1966; Conrad, 1963, 1964). However, these effects may well be memory and load dependent; certainly in such studies ample time was always available, and the stimuli were frequently single letters. A similar comment may be made with respect to the visual buffer store studies of Posner, Boies, Eichelman, and Taylor (1969) and Sternberg (1969). Rubenstein (Rubenstein, Garfield, & Millikan, 1970; Rubenstein, Lewis, & Rubenstein, 1971) found that homophonic nonsense words (BRANE, SPAID, etc.) require longer decision times as to whether they are real words than nonhomophonic nonsense words (BRATE, SPOID), perhaps because the former excite more false matches requiring rejection. Other support for the inevitability of phonological recoding comes from Dick (1971)—naming precedes categorization, from Klapp (1971)—implicit speech cannot be inhibited even in a same–different matching task where it is not required, and from Dalrymple-Alford (1972)—it is difficult to suppress the reading responses in a Stroop task.

On the other hand, proponents of the direct access theory (cf. Bower, 1970; Kolers, 1970; Smith, 1971) would claim that phonological representation is not a necessary part of the

recognition of a visually presented word, perhaps being superfluous in the case of practiced readers, familiar material, and high reading speeds.

From the studies reported here, it does seem that actual and overt articulation of both items is necessary to produce the interfering effect with heterophones. Thus an articulatory set may be briefly established, which is briefly linked to a particular configuration of letters—a "grapheme-phoneme correspondence" (Gibson, Pick, Osser, & Hammond, 1962).

REFERENCES

Baddeley, A. D. Short term memory for word sequences as a function of acoustic semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 362–365.

Bower, T. G. R. Reading by eye. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic, 1970.

Conrad, R. Acoustic confusions and memory span for words. *Nature*, 1963, **197**, 1029–1030.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

Dalrymple-Alford, E. C. Sound similarity and color-word interference in the Stroop task. *Psychonomic Science*, 1972, **28**, 209–210.

Dick, A. O. Processing time for naming and categorization of letters and numbers. *Perception & Psychophysics*, 1971, **9**, 350–352.

Gibson, E. J., Pick, A., Osser, H., & Hammond, M The role of grapheme-phoneme correspondence i the perception of words. *American Journal o Psychology*, 1962, **75**, 554–570.

Klapp, S. T. Implicit speech inferred from respon latencies in same-different decisions. *Journal Experimental Psychology*, 1971, **91**, 262–267

Kolers, P. A. Three stages of reading. In I Levin & J. P. Williams (Eds.), *Basic studi reading*. New York: Basic, 1970.

Kolers, P. A., & Lewis, C. L. Bounding of lett sequences and the integration of visually presente words. *Acta Psychologica*, 1972, **36**, 112–124

Morton, J. Interaction of information in recognition. *Psychological Review*, 1969, **76** 165–178.

Posner, M. I., Boies, S. J., Eichelman, W & Taylor, R. L. Retention of visual and n codes of single letters. *Journal of Experim Psychology Monograph*, 1969, **79**(1, Pt. 2).

Rubenstein, H., Garfield, L., & Millikan, J Homographic entries in the internal lex *Journal of Verbal Learning and Verbal Beha* 1970, **9**, 487–494.

Rubenstein, H., Lewis, S. S., & Rubenstein, A. Evidence for phonemic recoding in vis word recognition. *Journal of Verbal Learning a Verbal Behavior*, 1971, **10**, 645–657.

Smith, F. *Understanding reading: A psycholinguistic analysis of reading and learning to read*. New York: Holt, 1971.

Sternberg, S. Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, 1969, **57**, 421–457.

(Received March 29, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 95–100

DIFFERENTIAL CLASSICAL CONDITIONING OF POSITIVE AND NEGATIVE SKIN POTENTIALS

KATHLEEN GLAUS[1] AND HARRY KOTSES

Ohio University

The positive and negative components of the skin potential response (SPR) were investigated in a differential classical conditioning paradigm employing a 10-sec. interstimulus interval (ISI). The *S*s were 30 undergraduate females. Analyses of variance indicated that conditioned effects, measured in terms of response frequency, were present in both SPR components, and that these occurred in early as well as late portions of the ISI. Intrasubject response comparisons revealed that *S*s exhibited a predominant response polarity and displayed both conditioned effects and orienting behavior in terms of that polarity. Procedural considerations in SPR conditioning were also discussed.

In classical conditioning of electrodermal responses, a great deal of attention has been devoted to skin resistance while only a few studies have been concerned with skin potential. Additionally, the work dealing with the latter response has largely been characterized by a lack of conditioned effects, possibly resulting from inappropriate control and scoring procedures. Loveless and Thetford (1966) reported a failure to observe skin potential response (SPR) conditioning. However, their design, incorporating test trials interspersed among conditioning trials in a single group, may have precluded a meaningful assessment of conditioned effects. Similarly, Yamazaki, Watanabe, and Niimi (1969) employed test trials during a series of simultaneous conditioning trials and evaluated conditioning in terms of the relative amplitudes of positive and negative SPR components on test trials. They concluded that since the negative responses increased across test trials while the positive component responses decreased, the negative component was more likely to reflect a conditioned effect. While interesting, this interpretation is difficult to sustain since the design employed lacks an appropriate conditioning control (Rescorla, 1967). Several studies have adequately dealt with the control problem but have failed to produce unequivocal results. Kotses (1969), in a study employing unpaired control groups where simultaneous recordings of skin potential and skin resistance were made, found no evidence of conditioning in either SPR component, although the skin resistance response (SRR) showed a strong conditioned effect. In this case, SPR conditioning may have been obscured by the independent groups design, by the scoring system employed,[2] or possibly by the fact that pairing effects were evaluated in terms of response amplitude. Finally, successful conditioning of the positive potential, in terms of response magnitude, was reported by Shmavonian, Miller, and Cohen (1968), who utilized a differential conditioning paradigm in a study designed to examine age and sex differences in autonomic conditioning. Differential responding was demonstrated only in the young *S*s (ages 17–25), and no evidence for negative potential conditioning was obtained. The lack of more general conditioned effects in this study may have been due to attenuated overall responsiveness resulting from the presentation of a series of adaptation trials prior to the acquisition session.

The predominant use of the SRR and subsequent neglect of the SPR in electrodermal conditioning have been due largely

[1] Requests for reprints should be sent to Kathleen Glaus, Department of Psychology, Ohio University, Athens, Ohio 45701.

[2] Negative component amplitudes were scored as the sum of all negative potentials following onset of the conditioned stimulus. These were measured between negative response onset and peak negative level. Positive component amplitudes were measured between peak negative and peak positive levels. Only one positive deflection but as many as 2 negative deflections were recorded per trial.

to problems encountered in the evaluation and interpretation of the positive, negative, and multiphasic potential wave forms (Holmquest & Edelberg, 1964). While the SRR does offer certain advantages in terms of simplified scoring and interpretation, these may be outweighed by distortions resulting simply from the passage of an electric current through the skin which may act in some way to alter skin structures. The SPR is inherently free of this source of artifact and its complex wave forms may offer a further advantage in yielding ad-ditional information not available in the uniphasic SRR (Edelberg, 1967). With regard to the latter, a number of investigators have suggested that the component wave forms of the SPR may be differentially sensitive to conditioning (Shmavonian et al., 1968; Yamazaki et al., 1969) and to various perceptual and cognitive factors of psychological interest (Burstein, Fenz, Bergeron, & Epstein, 1965; Forbes, 1936; Forbes & Bolles, 1936; Hupka & Levinger, 1967; Raskin, Kotses, & Bever, 1969). Thus, of the 2 electrodermal responses, the SPR is potentially a less ambiguous measure than the SRR and one which may yield additional information concerning complex variables in classical conditioning.

The purpose of the present study was twofold. First, an attempt was made to determine which components of the SPR are modified in classical conditioning. Thus the relative effect of conditioning on the frequency of positive and negative potentials was examined. Second, the differential effect of conditioning on long and short latency responses was assessed. This aspect was of interest in light of inconsistent evidence from SRR studies dealing with the temporal locus of conditioned effects in a long interstimulus interval (ISI). While it is a general finding that responses appearing late in a long ISI are reliably affected by pairing (Gale & Ax, 1968; Gale & Stern, 1967; Leonard & Winokur, 1963; McDonald & Johnson, 1965; Prokasy & Ebel, 1967; Stewart, Stern, Winokur, & Fredman, 1961), some studies have failed to demonstrate conditioned effects for early responses (Leonard & Winokur, 1963; McDonald & Johnson, 1965; Stewart et a[illegible] 1961). This problem appeared to be [illegible] some significance with respect to SPRs be[illegible] cause of the possibility that potentials [illegible] different polarity would be associated wi[illegible] early and late portions of the ISI. Finall[illegible] an attempt was made in the present stu[illegible] to avoid certain of the problems obser[illegible] in previous SPR conditioning work. Sp[illegible] ically, a differential conditioning paradi[illegible] not preceded by an adaptation session w[illegible] employed to provide a precise intrasubj[illegible] control for sensitization and to elimin[illegible] the uncertain effects of prior habitua[illegible] Additionally, response frequencies, ra[illegible] than magnitude measures, were obtai[illegible]d since the former would seemingly be l[illegible] sensitive to artifact introduced by int[illegible] actions between subsequent positive a[illegible] negative wave forms (Lykken, Miller, & Strahan, 1968).

Method

Subjects. Thirty female undergraduate psychology students, ranging in age from 18 to 23 [illegible] served as *Ss*. Subjects were volunteers and received extra course credit for participation.

Apparatus. Skin potentials were picked up by 2 Beckman Ag–AgCl electrodes having a diameter of 1.5 cm., amplified by a Grass model 7P1B low-level dc preamplifier, and displayed on one channel of a Grass model 7B polygraph operated at a paper speed of 25 mm/sec. The electrodes were attached to the skin with Beckman electrode collars, and Beckman electrode paste served as the electrolyte. Lights serving as conditioned stimuli (CSs) consisted of 2 7.5-w. white lights having a duration of 10 sec. These were mounted side by side in 5 × 5 × 10 cm. individual metal compartments which were completely enclosed with the exception of the side facing *S*. This side was covered by 2 5 × 5 cm. metal slides whose centers consisted of an open circle 4.5 cm. in diameter. Red and yellow sheets of translucent plastic were placed over the circles to diffuse the light and provide chromatic differences between the 2 signals. These stimuli were presented at eye level approximately 5 ft. in front of the seated *S*. The unconditioned stimulus, a 500-msec. 112-db. white noise, was produced by a Grason-Stadler model 455B white noise generator, amplified by a locally constructed audio amplifier and presented to *S* via Koss PRO-4 headphones. Onset and duration of both visual and auditory stimuli were controlled by BRS-Foringer solid-state control equipment. During recording, *S* was seated in a comfortable armchair located in a 3 × 5 m. room with an ambient sound level of 45 db. All sound

levels were measured with a General Radio sound-level meter, model 1565-A.

Procedure. A differential delay classical conditioning paradigm with a 10-sec. ISI was employed. Twenty CS+ and 20 CS− trials, wherein CS color was counterbalanced across *S*s, were presented in random order with the stipulation that no more than 3 trials of either type occur in succession. Intertrial intervals varied between 40 and 80 sec. around a mean of 60 sec.

Upon arrival at the laboratory, *S* was escorted into the experimental room where *E* applied the electrodes. Electrode sites were cleaned with a 70% solution of ethanol, and a Kimwipe was used to abrade the skin under the inactive electrode. The active electrode was positioned over the thenar eminence of the left hand and the reference was attached to the volar surface of the left forearm approximately 8 cm. below the elbow. A ground electrode was attached to the dorsal surface of the right hand. Standard instructions read to *S* stressed the importance of refraining from unnecessary movement and remaining awake and alert during the session. Following this, *E* positioned the headphones on *S* and entered the recording room to initiate the session.

Scoring of skin potentials. The ISI was divided into 2 scoring intervals, .4–5.0 sec. and 5.1–10.0 sec. after CS onset, and all skin potentials were classified as either first- or second-interval responses according to their onset latencies. Since large changes in the tonic skin potential level were rarely observed during the ISI, the base level at stimulus onset on each trial was employed as a reference from which the polarity of subsequent phasic responses was determined. Thus, uniphasic potentials were scored as negative or positive responses, respectively, according to whether pen deflection was in the direction of increasing or decreasing negativity relative to the reference level. When 2 or more responses occurred in succession (diphasic and polyphasic responses), additional criteria were employed. Specifically, if all phasic responses were in the region of increased negativity relative to the reference level, successive negative-going deflections were scored as negative responses while positive-going movements of the pen were judged as returns to base level. Similarly, for phasic responses in the region of decreased negativity, successive positive-going deflections were scored as positive responses and negative-going movements of the pen were seen as returns to base level. In cases where the peak-to-peak excursion of the pen crossed the reference level, successive peaks were scored, as were uniphasic responses, according to the polarity of the region in which each occurred.

Only deflections representing a .5-mv. or greater change in potential and having a slope of at least 10° were counted as responses. Artifact due to movement or to faulty electrical contact was eliminated on the basis of extremely high slopes (approaching 90°) and irregular wave shape.

The frequency measures obtained included all criterion responses occurring on each trial since, at present, no strong theoretical basis exists for restricting attention to any one response and since the total frequencies reflected the relative degree of positive and negative skin potential activity.

Results

Conditioned effects in positive and negative potentials were observed in both portions of the ISI. Separate 4-way repeated measures analyses of variance for the 2 wave forms, wherein intervals, pairing, and trials were considered fixed factors and *S*s a random factor, supported these observations. These analyses yielded reliable pairing effects for both positive and negative potentials, $F\ (1, 29) = 28.05$ and 43.34, respectively, $p < .001$. Furthermore, as shown in Figure 1, conditioning appeared to be more pronounced in the second interval for both positive and negative potentials. This observation received only partial statistical support, however, since a reliable Pairing × Intervals interaction was obtained only for the negative response, $F\ (1, 29) = 9.84$, $p < .005$. Additionally, it should be noted that tests of simple main effects nesting pairing within intervals for the negative response resulted in reliable pairing effects for the first, $F\ (1, 58) = 11.05$, $p < .005$, as well as the second, $F\ (1, 58) = 51.04$, $p < .001$, interval. Finally, reliable Pairing × Trials interactions were obtained for both positive, $F\ (19, 551) = 3.58$, $p < .001$, and negative, $F\ (19, 551) = 1.89$, $p < .01$, potentials, which appeared to reflect the typical curvilinear nature of autonomic conditioning functions.

The overall frequency of responding was higher in the first interval than the second for both positive, $F\ (1, 29) = 24.95$, $p < .001$, and negative, $F\ (1, 29) = 39.43$, $p < .001$, potentials. It was also apparent that the negative potential evidenced a degree of response decrement across trials in the first, $F\ (19, 551) = 2.34$, $p < .001$, but not in the second interval, despite a reliable overall trials effect, $F\ (19, 551) = 3.59$, $p < .001$. In the positive response, only a trials effect was obtained, $F\ (19, 551) = 6.66$, $p < .001$.

Examination of intrasubject response frequencies in both ISI intervals revealed

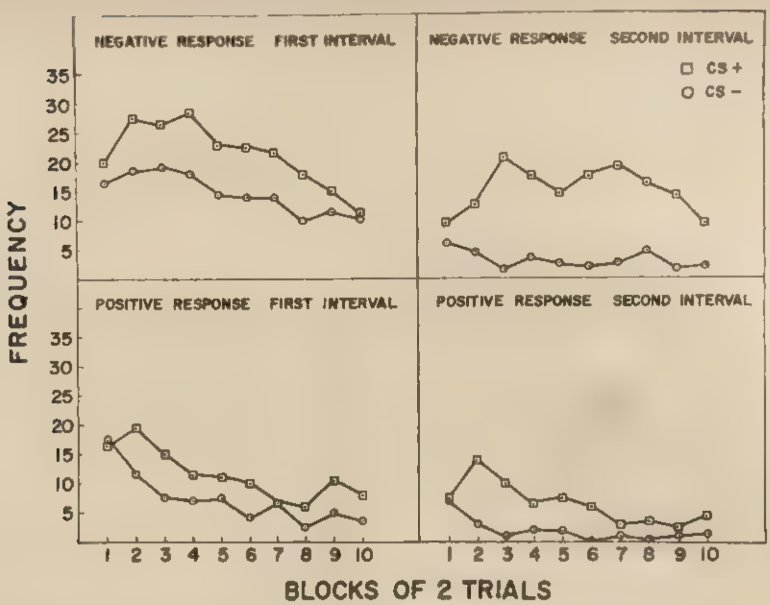


FIGURE 1. Frequency of positive and negative skin potential responses to positive and negative conditioned stimuli (CS+, CS−) in first and second portions of the interstimulus interval averaged over blocks of 2 trials.

additional details regarding the relative effect of conditioning on positive and negative potentials. These data indicated that in terms of overall frequency, *S*s tended to show a predominant response polarity. On the basis of the within-*S*s algebraic difference between frequency of responses to CS+ and CS−, [f(CS+) − f(CS−)], across the entire ISI, pairing effects were observed in 29 of the 30 *S*s, but these were greater across negative potentials in 20 *S*s and across positive potentials in 7 *S*s, while 2 *S*s failed to show a distinct pattern. Individual data also showed that in spite of overall conditioned response latencies, *S*s showing greatest conditioning across positive potentials exhibited such differences primarily during the first half of the ISI (5 of 7 *S*s), while those conditioning in the negative response showed greatest differentiation in the second interval (15 of 20 *S*s). A Fisher's exact probability test performed on these frequencies revealed that they were statistically significant, $p < .05$.

Since *S*s exhibited a strong tendency to condition primarily in terms of a single polarity, individual total responsiveness was examined to determine the degree to which each *S* responded in a consistent fashion to conditional and control stimuli. In 20 of the 30 *S*s, at least two thirds of the total responses made were of like polarity. Of the 20 *S*s showing such a degree of response predominance, 16 were negative responders while 4 responded positively.

Discussion

The results of the present study indicate that both components of the SPR are modified in classical conditioning, and further that conditioned effects occur in early as well as late portions of the ISI. Previous related research dealing with differential behavior of positive and negative potentials in classical conditioning has led to some inconsistent conclusions. Yamazaki et al. (1969) stated that the effects of conditioning are observed primarily in the negative potential, while Shmavonian et al. (1968) argued that the positive potential is most sensitive to pairing. The current research yielded no evidence that conditioning is manifested exclusively in either component. Although there appeared to be a greater pairing effect in terms of the negative component, this simply may have been a function of higher negative response frequency.

Dividing the ISI into 2 intervals and examining responses within these intervals separately permitted a determination of the temporal characteristics of the conditioned activity and provided some information with respect to the orienting response (OR). A comparison of short and long latency conditioned responses showed that conditioned effects for the negative potential are more pronounced in the latter portion of the ISI. This finding is in accord with the results of previous SRR studies employing long ISIs in that, in these studies as a whole, conditioned effects appeared more consistently later in the interval (Gale & Ax, 1968; Gale & Stern, 1967; Leonard & Winokur, 1963; McDonald & Johnson, 1965; Prokasy & Ebel, 1967; Stewart et al., 1961). It should be noted, however, that with the negative SPR and probably with the SRR, the attenuated pairing effects observed initially may result from OR components present primarily in the first portion of the interval. In the present study this was supported by the increased frequency of response to the CS− early in the ISI for the negative response. Elevated CS− response frequencies were also observed in the first interval for the positive potential; however, in this case, attenuated early pairing effects did not result.

Related to the observations concerning response frequency to the unpaired stimuli are the findings of investigators who have addressed themselves directly to the relationship between SPR polarity and orienting behavior. As with the conditioned response, some have associated orienting with the negative component (Raskin et al., 1969; Shmavonian et al., 1968), while others have assigned this property to the positive component (Loveless & Thetford, 1966; Yamazaki et al., 1969). In the present study, elevated frequencies were observed in the first interval for both potentials, suggesting that the OR may be manifested in either component.

Intrasubject response comparisons permitted a more precise interpretation of the effects of pairing on positive and negative skin potential components. Most *S*s exhibited a degree of differential responding in terms of both components. However, there was a tendency for *S*s to condition predominantly in terms of one component or the other. This propensity may reflect a stable characteristic of individual responsiveness, but probably can be better accounted for in terms of the interaction between individual and situational response determinants. The latter possibility is more likely in view of studies relating SPR components to stimuli of varying levels of threat, apprehension, or intensity (Burstein et al., 1965; Forbes & Bolles, 1936; Raskin et al., 1969). The previous work dealing with such relationships has been limited somewhat by its failure to consider individual differences in response to the stimulus. A model of SPR polarity incorporating individual response differences of this nature appears to be required in the interpretation of the results of the present study, since the same stimuli elicited positive and negative SPR components to varying degrees in different *S*s.

An additional finding apparent in the data of individual *S*s concerned the principal locus of differential responding for individuals exhibiting a predominant response polarity. Subjects showing a conditioned effect primarily in terms of the negative component displayed greater differential responding in the latter half of the ISI, while those showing more positive component conditioning did so earlier. This effect, which may in part account for the lack of differential pairing effects between the first and second intervals in the grouped data for the positive response, was probably due to the habituation rates of responses to the CS−, which appeared to be greater for the positive than for the negative potential in the first interval. While the reason for differential rates of habituation of positive and negative components is unclear, the effect has been noted by other investigators (Loveless & Thetford, 1966; Raskin et al., 1969; Yamazaki et al., 1969) who have, partially on the basis of such differences, attempted to relate response polarity to orienting behavior. Such an interpretation, of course, would not be consistent with a model of SPR activity emphasizing individual as well as situational determinants of responsiveness.

In summary, on the basis of the current research, several conclusions may be advanced tentatively regarding SPR conditioning which also relate to procedural considerations. First, since *S*s exhibit a predominant response polarity and have a tendency to display conditioned effects as well as orienting behavior primarily in terms of that polarity, a differential paradigm which provides intrasubject control for sensitization appears to be required. Second, because of the differential habituation rates of the positive and negative components, an adaptation session prior to conditioning may complicate matters and probably should be avoided. Finally, comparisons between the

present study and previous work suggest that frequency measures may provide a better index of conditioning since such measures presumably would be less distorted by rapid reversals in response polarity than would amplitude measures. Of course, further research is required to provide more definitive information concerning these issues.

REFERENCES

Burstein, K., Fenz, W., Bergeron, J., & Epstein, S. A comparison of skin potential and skin resistance responses as measures of emotional responsivity. *Psychophysiology*, 1965, **2**, 14–24.

Edelberg, R. Electrical properties of the skin. In C. C. Brown (Ed.), *Methods in psychophysiology*. Baltimore: Williams & Wilkins, 1967.

Forbes, T. W. Skin potential and impedance measures with recurring shock stimulation. *American Journal of Physiology*, 1936, **117**, 189–199.

Forbes, T. W., & Bolles, M. M. Correlation of the response potentials of the skin with "exciting" and "non-exciting" stimuli. *Journal of Psychology*, 1936, **2**, 273–285.

Gale, E. N., & Ax, A. F. Long-term conditioning of orienting responses. *Psychophysiology*, 1968, **5**, 308–315.

Gale, E. N., & Stern, J. A. Conditioning of the electrodermal orienting response. *Psychophysiology*, 1967, **3**, 291–301.

Holmquest, D., & Edelberg, R. Problems in the analysis of the endosomatic galvanic skin response. *Psychophysiology*, 1964, **1**, 48–54.

Hupka, R., & Levinger, G. Within subject correspondence between skin conductance and skin potential under conditions of activity and passivity. *Psychophysiology*, 1967, **4**, 161–167.

Kotses, H. Effects of orienting and defensive reflexes in classical conditioning. (Doctoral dissertation, Michigan State University) Ann [illegible]or, Mich.: University Microfilms, 1969. No. [illegible]-16, 154.

Leonard, C., & Winokur, G. Conditioning [illegible]rsus sensitization in the galvanic skin r[illegible]nse. *Journal of Comparative and Physiological* [illegible]*chol*[illegible], 1963, **56**, 169–170.

[illegible]ess, E., & Thetford, P. E. Int[illegible] and conditioning of the positive an[illegible] components of the skin potential r[illegible]nse. *Psychological Record*, 1966, **16**, 357–36[illegible]

Lykken, D. T., Miller, R. D., & Stra[illegible] F. Some properties of skin conductance [illegible] skin potential. *Psychophysiology*, 1968, **5**

McDonald, D. G., & Johnson, L. C. A[illegible]sis of GSR conditioning. *Psychophysiology*, [illegible]5, **1**, 291–295.

Prokasy, W. F., & Ebel, H. C. Three [illegible]inct components of the classically condition[illegible] GSR in human subjects. *Journal of Exp*[illegible]*mental Psychology*, 1967, **73**, 247–256.

Raskin, D. C., Kotses, H., & Bever, J. A[illegible]nic indicators of orienting and defensive r[illegible]flexes. *Journal of Experimental Psychology*, 1[illegible], **80**, 423–433.

Rescorla, R. Pavlovian conditioning and it[illegible] proper control procedures. *Psychological Review* 1967, **74**, 71–80.

Shmavonian, B. M., Miller, L. H., & Coh[illegible]n, S. I. Differences among age and sex groups in electrodermal conditioning. *Psychophysiology*, 1968, **5**, 119–131.

Stewart, M. A., Stern, J. A., Winkour, G., & Fredman, S. An analysis of GSR conditioning. *Psychological Review*, 1961, **68**, 60–67.

Yamazaki, K., Watanabe, T., & Niimi, Y. Orienting components of the skin potential reflex in classical conditioning. *Japanese Journal of Psychology*, 1969, **40**, 98–102.

(Received March 29, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 101–105

LEVELS OF PROCESSING IN WORD RECOGNITION AND SUBSEQUENT FREE RECALL[1]

JOHN M. GARDINER[2]

Birkbeck College, University of London, London, England

The present experiment explored a situation in which *S*s were unexpectedly required to recall the target words from a perceptual decision-making task. The targets were defined with respect either to their phonemic or semantic attributes, and *S*s held these attributes in "working memory" for varying time intervals prior to target presentation. Semantically defined targets were better recalled subsequently than were phonemically defined targets, although the latter gave rise to longer decision latencies in the initial task. Also, subsequent target recall was not affected by the length of time the target-defining attributes had been held in working memory. These results were discussed within the context of Craik and Lockhart's "levels-of-processing" approach.

Craik and Lockhart (1972) described a framework for memory research in which the memory trace is viewed essentially as the by-product of perceptual analyses. Central to their argument is the notion that the stability of the memory trace is a positive function of the type and depth of processing involved in the encoding of perceptual events. Greater depth of processing is defined in terms of the degree of cognitive involvement in carrying out stimulus analyses. Memory is assumed to be tied to a continuum of levels of processing which range, for example, from sensory analyses to the activation of associative semantic attributes.

As Craik and Lockhart (1972) suggested, this formulation implies that research should be directed toward determining the memorial consequences of various types of perceptual operations. Craik (in press, Experiments IV and V), for example, reported 2 studies in which *S* was given an initial perceptual decision-making task followed by an unexpected memory test. The purpose of the initial decision-making task was to lead *S* to encode different words at different levels of analysis. For each of a series of words presented briefly in a tachistoscope, *S* was asked a question such as (*a*) Is there a word present? (*b*) Does the word rhyme with ____? (*c*) Is the word a member of the following category? or (*d*) Does the word fit into the following sentence? In general, Craik's results showed that "deeper" decisions about words gave rise both to longer decision latencies in the initial task, and to better memory performance subsequently.

These results are in good agreement with the levels-of-processing view. However, it could be argued that the differences obtained in memory performance were due to the corresponding differences in initial processing time, rather than due to increasing "depth" of analysis. Although words processed at "sentence level" were better remembered than words processed at "rhyme level," for example, *S* also took longer in making sentence level decisions.

The general aim of the present experiment was to obtain further evidence in support of the levels-of-processing approach. Again, the strategy was to examine the effects of depth of analysis in a perceptual decision-making task upon subsequent memory for "target" words. The present study was also designed to provide more information on the effects of initial processing time in such a situation. In this context, it is important to note that processing time refers, poten-

[1] This research was supported by a Science Research Council Postgraduate Award. The author is grateful to Fergus I. M. Craik and Peter Herriot for their helpful comments on an earlier draft of this article.

[2] Requests for reprints should be sent to John M. Gardiner, who is now at the Department of Social Science and Humanities, The City University, St. John Street, London EC1V 4PB, England.

tially, both to the time *S* takes to process the target word as such and to the time *S* spends in processing those attributes that serve to define the target. The present paradigm permits the effects of these 2 processing times to be evaluated separately.

In essence, *S*s were presented with a series of word sets, each containing a unique target word. Prior to the presentation of each word set, *S* was given the "defining attributes" of the target word. For half of the *S*s, the target words were defined by their containing either 1 or 2 phonemes. For the remaining *S*s, the targets were defined as members of either a single category or a nested category. The target word POLAND, for example, might be defined as containing the phonemes /l/ or the phonemes /l/ and /n/, on the one hand, or as a country or a European country, on the other hand. Thus, one group of *S*s was required to identify targets at a "structural" level of analysis while the other group of *S*s had to identify targets at a deeper, semantic level. The *S*'s decision latencies in response to targets provided a measure of target-word processing time. In order to examine the effects of processing time for target-defining attributes, the number of nontarget words preceding the target in each set varied across trials. Thus, the length of time the target-defining attributes were held in "working memory" was indexed by the position of the target in any given word set. Following this perceptual decision-making task, *S*s were unexpectedly asked to write down all the target words they could remember.

The experiment had, therefore, the following 2 main objectives: (*a*) to compare recall performance following either phonemic or semantic processing of target words, and (*b*) to examine the relationship between initial processing times and subsequent recall.

Method

Subjects. The *S*s were 32 volunteers, mostly undergraduate students, who were either paid for their services or were fulfilling a course requirement. The *S*s were tested individually.

Design. The experiment was conducted [illegible]ing a mixed design. Level of processing (pho[illegible]r semantic) was the between-*S*s variable, ca[illegible]re for convenience "attribute type." Sixteen [illegible]re assigned at random to each between-*S*s [illegible]n. Within *S*s, the design comprised a factorial [illegible]om-bination involving 2 levels of "attribute nu[illegible]ber" and 4 levels of "attribute lag." Attribute nu[illegible]ber refers to target-word definition by either 1 [illegible]r 2 phonemes, on the one hand, or by a si[illegible] or nested category name, on the other hand. At[illegible]ute lag refers to the position of the target with[illegible]ach word set.

The *S*s were presented with 38 word [illegible]he target-recognition session. Prior to th[illegible]ta-tion of the first word in each set, th[illegible]for that set was defined either as a member o[illegible]gle or nested category for *S*s in the semant[illegible]di-tion, or by the word containing either 1 o[illegible]ven phonemes for *S*s in the phonemic conditio[illegible]The target word was always the final word in [illegible]set. On any given trial, it was preceded by 1 [illegible]n-target words. The critical positions of th[illegible]get for scoring purposes were 3, 4, 5, and 6. [illegible]ons 2 and 7 were included to reduce target pre[illegible]bil-ity by position; 6 of the 38 trials were of t[illegible]nd. The remaining 32 target words appeared [illegible]ally often at each of the 4 critical positions an[illegible] were defined equally often by either 1 or 2 att[illegible]butes in each between-*S*s condition. For presentation purposes, the total number of word sets was divided into 4 blocks which were counterbalanced with regard to presentation order and the level of attribute number. In effect, all *S*s received the same target words, and no target word was differentially favored with respect to order of presentation or experimental condition.

Approximately 2 min. after the final trial in the recognition sessions, *S*s were unexpectedly asked to write down, in any order, all of the target words they could remember.

Materials and procedure. Of the 32 critical target words, 27 were selected from the Battig and Montague (1969) category norms and the remaining 5 were chosen by *E*. Those drawn from the published norms had a mean frequency of occurrence of 16%. This relatively low value was chosen in order to minimize *S*'s chances of guessing the target word when given category names as the target-defining attributes. The frequency count for nontarget words ranged between 5–20/million (Thorndike & Lorge, 1944). All words, both targets and nontargets, were 2 syllables in length.

Two sets of presentation lists were constructed for each of the 4 levels of attribute lag. The difference between the 2 sets lay in the nontarget words in each; specifically, where target words were defined by 2 attributes, the preceding nontarget words always included 1–3 words that contained one or the other of the 2 defining phonemes and belonged to the single but not to the nested category. This procedure was designed to ensure that *S*s had to utilize both given attributes when targets were so defined.

The word lists were typed in uppercase letters for presentation via an *IBM* electric typewriter and a closed-circuit television system. A Birkbeck Laboratory Timer and Signal Source was used to both trigger the presentation of each word and also to start, simultaneously, a Venner millisecond stopclock. The Birkbeck Laboratory Timer and Signal Source was operated manually by *E*, contingent upon *S*'s having responded to the word currently presented. In practice, words within a set appeared at about a 2-sec. rate.

The *S*'s vocal response, amplified from a throat microphone, stopped the Venner timer. Each *S* was instructed to respond *yes* or *no* to each word as rapidly but as accurately as he could. In addition, *Ss* were instructed that after each *yes* response they were to report the target word aloud. The decision latency measure, however, was of course from the onset of the target word to *S*'s *yes* response.

The *Ss* were initially familiarized with the apparatus and procedure. They were also fully instructed as to the nature of the stimulus material. Prior to the experimental trials, each *S* received 6 practice trials in which the target word appeared once at each possible target position. Each trial began with *E* reading aloud the target-defining attributes for the word set. Consider, for example, the case where the target word was POLAND. If the target was defined by one attribute, *E* said either "the target contains the sound /l/," or "the target is a country," for *Ss* in the phonemic and semantic conditions, respectively. Where the target was defined by 2 attributes, *E* said either "the target contains the sounds /l/ and /n/," or "the target is a European country." Examples of target words and their corresponding 1- or 2-attribute definitions were: LEOPARD: /p/, or an animal; /p/ and /d/, or a wild animal; DUBLIN: /b/, or a city; /b/ and /n/, or a capital city; RICKETS: /k/, or a disease; /k/ and /t/, or a bone disease; TEMPLE: /m/, or a building; /m/ and /p/, or a religious building. Following target definition, each word in the set was presented, one word at a time. Examples of 2 presentation sets were: FATHOM, METEOR, ATTIC, POLAND, and FATHOM, JAPAN, ATTIC, POLAND. The latter example includes one distractor which contains 1 of the 2 target-defining phonemes and belongs to the main but not to the nested category: sets including one or more such distractors were presented when the target was defined by 2 attributes. Each trial terminated after *S* had correctly identified and named the target; the next trial began immediately.

The *Ss* were told that the point of the experiment was to investigate the effects of defining target words by 1 or 2 "values along a dimension" on the speed with which they could identify the target. No *S* reported subsequently that he had anticipated having to recall the target words.

Unknown to *Ss*, only decision latency in response to target words was recorded. Errors made by *S* in the target-recognition session were noted and "corrected" orally by *E* at the time of their occurrence.

RESULTS

For both the decision latency data and the recall data, Target Positions 3, 4, 5, and 6 (attribute lag) were collapsed into 2 levels of 3 plus 4 (short lag) and 5 plus 6 (long lag) in order to obtain more reliable estimates. The *E* errors and apparatus failures accounted for less than 2% of all responses in the target-recognition session; *S* errors accounted for 5% of responses in the semantic condition and 6% in the phonemic condition. Error data were excluded from all analyses. The results are shown in Figure 1.

Consideration is given first to the latency data. The basic datum was median reaction time for correct responses from the 8 trials each *S* received in the $2 \times 2 \times 2$ combination representing attribute type (phonemic or semantic), attribute number (1 or 2), and attribute lag (short or long). The mean median scores are shown in the left-hand panel of the figure; each point is based on approximately 120 observations. These data were subjected to an analysis of variance. The main effects due to attribute type, $F(1, 30) = 14.9$, attribute number, $F(1, 30) = 43.5$, and attribute lag, $F(1, 30) = 8.8$, were all significant at or beyond $p < .01$. The only significant interactions were the following: Attribute Type × Number, $F(1, 30) = 12.3$, $p < .01$, and Attribute Type × Lag, $F(1, 30) = 4.4$, $p < .05$.

No detailed interpretation of these results will be attempted. From the point of view of the subsequent argument, the most relevant feature in these data is that *Ss* took considerably longer to process target words at the phonemic level than they did when targets were defined semantically. This result might seem surprising in view of Shulman's (1970) finding that, in a probe-recognition task, homonyms were recognized more rapidly than synonyms. However, it should be noted that whereas homonyms can be identified on a "holistic" basis, target words in the present paradigm were identified on the basis of individual, constituent phonemes.

The right-hand panel in Figure 1 shows the percentage of target words which *Ss*

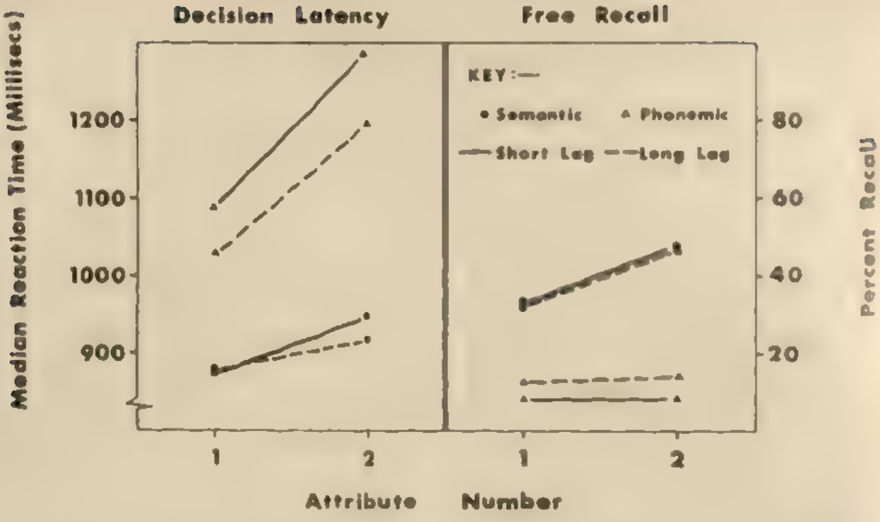


FIGURE 1. Decision latency and subsequent recall as a function to attribute type, number, and lag.

recalled in the final test, with the lag variable again collapsed into short and long lag. In order to obtain "workable" numbers for the purposes of analysis, the recall data supplied by each consecutively tested pair of *S*s were combined to form 8 macro-*S*s in each between-*S*s condition. An analysis of variance carried out on the macro-*S* recall scores confirmed that the main effect due to attribute type was highly significant, F (1, 14) = 73.2, $p < .001$. The main effect due to attribute number was also significant, F (1, 14) = 7.2, $p < .05$, but the main effect due to attribute lag was not ($F < 1$). The only significant interaction was Attribute Type × Number, F (1, 14) = 5.7, $p < .05$.

Not surprisingly, these results confirm that target words were recalled much better following semantic processing, as compared with phonemic processing, in the initial session. Recall of targets following phonemic processing is virtually at floor level. The results also indicate that the number of phonemes involved in processing the target had no effect upon recall performance. This finding contrasts with the comparable result in the semantic condition, where recall is better following target definition by a nested category than by a single category name.

Finally, it is apparent that the length of time the target-defining attributes were held in working memory (short or long attribute lag) did not affect subsequent recall of the target word. Neither the main effect due to attribute lag nor any interaction term involving lag approached significance.

Discussion

These results have several implications for the levels-of-processing approach discussed earlier. First, it is clear that processing target words at a semantic level led to good subsequent recall of the targets, whereas targets processed at a phonemic level were very poorly recalled. Similar differences in memory performance following semantic as opposed to structural processing have been reported by Craik (1973), Hyde and Jenkins (1969), and Schulman (1970). Such findings are in good agreement with the levels-of-processing view.

Although further analysis at a semantic level resulted in a corresponding improvement in recall (nested vs. single category names), additional analysis at a phonemic level did pot benefit recall. Thus, as Craik (1973) nointed out, the persistence of the memory trace should not be conceived to be simply a function of the number of analyses performed on the stimulus, but rather, a function of the degree of meaningfulness extracted from it.

The markedly poor recall of target words that had been processed at a phonemic level was associated with much longer decision latencies for these targets in the orienting task. Clearly, the pattern of results obtained in the recall test cannot be attributed to corresponding differences in the processing time for target words. It should be noted that following target-word recognition, *S*s were required to report the target word aloud. This procedure ensured that all target words were processed to at least a "nominal" level. The poor recall of phonemically defined targets must reflect the manner in which the words were processed, rather than the possibility that the target word as a whole was never perceived.

It was also found that the length of time the target-defining attributes were held in working memory did not affect subsequent recall of the target words. This finding relates well to Craik and Lockhart's (1972) suggestion that maintaining information at a "fixed" level of processing does not necessarily lead to any improvement in the memory trace. More generally, the present data demonstrate that duration at encoding is not nearly so crucial for registration in memory as are the levels of analysis involved. No evidence was found to support the notion that processing time, whether for target words or for target-defining attributes, was a critical factor for subsequent recall performance. The results imply that processing time may only be important insofar as it permits further or deeper levels of analysis to be carried out.

In summary, the present study explored a situation in which *S* was unexpectedly required to recall the target words he had identified in a perceptual decision-making task. The main findings were, first, recall of the target words was better following semantic as opposed to phonemic analysis of the target. Second, whereas further analyses at a semantic level led to a further improvement in recall, additional analyses at a phonemic level did not benefit recall. Third, neither processing time for target words, nor the length of time for which the target-defining attributes were held in working memory, appeared to be critical factors in determining subsequent recall performance. These results provide further support for the levels-of-processing view (Craik & Lockhart, 1972). It was concluded that the memorial consequences of the operations carried out in the perceptual decision-making task were attributable solely to the level of analysis these operations involved.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt. 2).

Craik, F. I. M. A "levels of analysis" view of memory. In P. Pliner, L. Krames, & T. M. Alloway (Eds.), *Communication and affect: Language and thought*. New York: Academic Press, 1973.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671–684.

Hyde, T. S., & Jenkins, J. J. Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, 1969, **82**, 472–481.

Schulman, A. I. Recognition memory for targets from a scanned word list. *British Journal of Psychology*, 1970, **62**, 335–346.

Shulman, H. G. Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 499–508.

Thorndike, E. L., & Lorge, I. *The teacher's wordbook of 30,000 words*. New York: Columbia University, Teachers College, Bureau of Publications, 1944.

(Received April 9, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 106–113

VISUAL AND AUDITORY CODING IN A MEMORY MATCHING TASK [1]

LARRY E. WOOD [2]

University of Wisconsin—Parkside

Two experiments were conducted to investigate modality-specific coding (auditory and visual) and recoding in a *same-different* reaction time task. On each trial 2 letters of the alphabet were presented sequentially in the same or in different modalities with an interstimulus interval of 150 msec or 3 sec. The coding dimension was inferred from differences in reaction time to contrasting sets of letter pairs, which were either aurally similar and visually dissimilar or vice versa. Similarity interacted with modality of the second stimulus, indicating that the stimuli were matched on the modality dimension of the second stimulus regardless of the modality of the first stimulus or the interstimulus interval. The results suggest that the first stimulus was recoded to the modality of the second for matching when the two were presented in different modalities and that visual coding was as effective as auditory coding in the task.

Although early research seemed to indicate that coding in short-term memory (STM) was auditory or acoustic in nature (e.g., Conrad, 1964; Wickelgren, 1965), later studies (e.g., Parks, Kroll, Salzberg, & Parkinson, 1972; Shulman, 1970; Tversky, 1969) have demonstrated considerable flexibility in coding processes and strategies. A task that has been particularly useful in studying coding processes is one in which *S*s are presented with 2 stimuli simultaneously or sequentially and are instructed to indicate as quickly as possible whether the stimuli are the same or different. Reaction time (RT) is the dependent measure of performance. For example, it has been found by several investigators (Beller, 1971; Posner, Boies, Eichelman, & Taylor, 1969; Posner & Mitchell, 1967) in a variety of conditions that *S*s are able to respond faster to pairs of letters that are physically the same (e.g., AA) than to pairs that only have the same name (e.g., Aa). The results seem to indicate that physical matches are made on the basis of visually coded information, but that name matches are made on the basis of aurally coded information. That is, before a name match is made, *S*s apparently recode the visual stimuli to their auditory equivalents, which leads to the longer RTs. More direct evidence that name matches are based on auditory information has been provided by Dainoff (1970) and Dainoff and Haber (1970), who found that name matches are affected by auditory similarity so that RTs were longer for similar pairs than for dissimilar pairs.

Tversky (1969) used a *same–different* RT task to demonstrate that the same information can be coded differently (either visually or verbally) depending on how *S* expects to use it. The stimuli consisted of line drawings of faces, each of which was associated with a name. Both types of stimuli varied on 3 binary dimensions, and the names and faces were paired so that similarity inversely covaried across the 2 types of stimuli. That is, faces differing by only one dimension were given names differing by 3 dimensions and vice versa, so that similar faces had dissimilar names and dissimilar faces had similar names. On each trial 2 stimuli were shown sequentially, with an interstimulus

[1] This study is based on a thesis submitted to the University of Iowa in 1971 in partial fulfillment of the requirements for the PhD degree. The author wishes to express his gratitude to James V. Hinrichs, who served as the chairman of the doctoral committee. The research reported herein was conducted while the author held Predoctoral Fellowship 1-F01-MH47899-01 awarded by the National Institute of Mental Health.

[2] Requests for reprints should be sent to Larry E. Wood, University of Wisconsin Parkside, Kenosha, Wisconsin 53140.

interval of 1 sec., and *S*s were asked to decide as quickly as possible whether or not the 2 stimuli had the same name. In each condition the first stimulus (S_1) was always either a name or a face, and the second stimulus (S_2) was randomly varied so that one type was shown on 79% of the trials and the other type was shown on the remaining 21% of the trials. Tversky reasoned that if *S*s could use either visual or verbal coding with equal ease and could readily recode stimuli from one mode to the other, then they should tend to code S_1 in the mode in which S_2 was expected (most frequently presented) regardless of the mode in which S_1 was presented. Such a strategy would result in longer RTs for those trials on which S_2 was presented in the unexpected modality because one of the stimuli would have to be recoded to the mode of the other before a match could be made. The predicted results were in fact obtained. Furthermore, analysis of the similarity effects showed that RT varied reliably as a function of the similarity of the S_2 mode, which indicated that the stimuli were matched on the S_2 mode independently of the S_1 mode.

Using a design similar to that employed by Tversky (1969), the present investigation was concerned with a comparison of visual and auditory coding and recoding in STM using letters of the alphabet as stimuli. The first experiment was performed to demonstrate that stimulus similarity manipulations could be used to infer the coding dimension, and the second experiment was an attempt to compare the relative effectiveness of visual and auditory coding and recoding as a function of the interval between stimuli to be matched by *S*. In order to determine the dimension on which stimuli were coded, the assumption was made that if RT in a matching task differed as a function of similarity along a specified dimension, then the stimuli were coded and matched on that dimension. Since the auditory and visual dimensions were the concern of the present investigation, letter pairs were constructed which were either high in auditory similarity and low in visual similarity (HALV) such as ZC, or low in auditory similarity and high in visual similarity (LAHV) such as OC. The expectation was that when stimuli were matched on the auditory dimension, HALV pairs would yield longer RTs than contrasting LAHV pairs. However, the opposite relationship would obtain between the 2 types of pairs when matches were made on the visual dimension.

Experiment I

Since differences in RT as a function of similarity were crucial in making inferences about stimulus coding, the first experiment was designed to demonstrate that the similarity manipulations described earlier (sets of contrasting HALV and LAHV stimulus pairs) would yield reliable differences. First, similarity effects are necessary in order to establish that coding and matching are occurring on the auditory or visual dimensions. Second, if the coding dimension in any given set of conditions is to be inferred from a particular pattern of similarity effects, it is necessary to show that a different pattern of similarity effects will obtain with the same sets of stimulus pairs and the same *S*s under different values of another independent variable. As discussed below, the independent variable chosen for this purpose was presentation modality.

Previous research with matching tasks has demonstrated that visual similarity affects RT with visual presentation of both stimuli of a pair (Clement & Carpenter, 1970) and auditory similarity affects RT with auditory presentation of both stimuli (McInish & Tikofsky, 1969). If presentation modality is a critical factor in coding for a matching task, then presentation of the same sets of contrasting HALV and LAHV letter pairs should reflect auditory coding with auditory presentation and visual coding with visual presentation. That is, with auditory presentation of both stimuli of each pair, HALV pairs should produce a longer mean RT than LAHV pairs, but with visual presentation, HALV pairs should produce a shorter mean RT

than LAHV pairs. Furthermore, the similarity effects should hold for the same *Ss* serving in both modality conditions.

Method

Subjects. The *Ss* were 10 female students from the introductory psychology courses at the University of Iowa. They were tested individually in 2 45-min. sessions on 2 successive days.

Design. Similarity (HALV and LAHV) and presentation modality (auditory and visual) were manipulated in a 2 × 2 factorial design with all *Ss* serving in all conditions.

Materials. As previously discussed, the predicted results of the present research were dependent upon significant differences in mean RT between contrasting pairs of letters conforming to a HALV or LAHV similarity classification. Therefore an attempt was made to construct letter pairs for which an opposite dichotomous classification on both similarity dimensions could be made. Visual similarity was defined simply as the amount of overlap in area between 2 uppercase letters, and auditory similarity was defined in terms of common phonemes shared by the letter names. Details of the similarity definitions can be found in Wood (1971). Four sets of letter pairs were constructed so that the 2 pairs in each set contained a common second letter with different first letters making up the appropriate similarity classifications. The four sets of HALV and LAHV pairs were ZC and OC, ZG and QG, VB and RB, and TP and RP. Both pairs of each set were terminated with a common letter in order to equate for perceptual differences among individual letters as a function of frequency, modality, etc., so that differences in RT between pairs could unambiguously be attributed to the similarity manipulations.

It should be noted that 6 of the 8 test pairs had a different first letter, which would enable *S* to predict the second letter of a pair rather accurately after being presented with the first letter. Hence, in order to create greater uncertainty regarding S_2 for the *different* pairs, 24 additional *different* pairs were constructed. Twelve of those pairs were constructed by combining each of the 6 different initial letters from the test pairs (Z, V, T, O, Q, R) with 2 of the 4 terminal letters (C, G, B, P) with which it had not been combined previously. The only restriction was that each of the 4 terminal letters was used for 3 of the pairs. The other 12 *different* pairs were constructed by pairing each of the 6 initial test-pair letters with the letters A and S as second letters. Finally, to create response uncertainty in the task, *same* pairs were constructed by combining each initial test-pair letter with itself 3 times, except for ZZ and RR which were used 4 times, making a total of 20 *same* pairs. Lists of 60 stimulus pairs were then constructed in which the 8 test pairs occurred twice each, the 24 additional *different* pairs occurred once each, and the 20 *same* pairs occurred once each.

Apparatus. The stimuli were presented visually on an Industrial Electronic Engineers Binaview display. For auditory presentation, the stimuli were tape-recorded and presented to *S* through earphones. The Binaview display was programmed by means of a punched-paper-tape reader which, in turn, was controlled by a BRS Electronics digital logic control system. The output of the tape recorder to be used for aural presentation was connected to the earphones through a set of relay contacts and to the BRS system for controlling the interstimulus interval (ISI) and intertrial interval for both auditory and visual presentation. A control signal from the paper-tape reader turned off the Binaview and closed the relay contacts to the earphones for aural presentation of a stimulus or turned on the Binaview and opened the relay contacts for visual presentation. Duration of the visual stimuli was 360 msec., the mean duration of the auditory stimuli as measured on an oscilloscope. The *Ss* responded to the stimuli by pushing one of 2 telegraph keys marked "same" or "different" and spaced 12 in. apart on a horizontal plane. The RTs (measured to the nearest millisecond) and the responses were recorded and punched onto paper tape.

Procedure. Each *S* received 5 blocks of trials, each consisting of a random order of the 60 stimulus pairs plus 6 practice pairs at the beginning of each block. The members of each stimulus pair were separated by 1 sec. measured from onset of the first stimulus (S_1) to onset of the second stimulus (S_2) and the pairs were separated by 3 sec. The order of the modality conditions was counterbalanced such that 5 of the *Ss* received visual presentation in the first session and 5 received auditory presentation first.

The *S* rested the index finger of each hand on one of the keys and was instructed to make a correct response as quickly as possible on each trial, indicating whether S_1 and S_2 were the same or different letters. A 2-min. rest period was given between each block of trials. The *same* response was always assigned to the dominant hand since there was no interest in directly comparing *same* and *different* responses.

Results

The first block of trials, the 6 practice trials in the remaining 4 blocks, and all trials on which errors were made were excluded from the analysis. Error rates ranged from 2%–5% across conditions. Since each test pair was presented twice in each block, there was a maximum of 8 correct trials per stimulus pair. A median RT was computed for each pair and then the 4 medians for each similarity condition (HALV or LAHV) were averaged, which resulted in a mean RT for

each S for each of the 4 similarity and modality conditions. These scores were then submitted to a 2-way analysis of variance.

The mean RT, *SE*, and percentage error for the 4 similarity and modality conditions are shown in Table 1. The hypothesis that similarity would have a different effect depending on the presentation modality was confirmed by a significant Similarity × Modality interaction, $F(1, 9) = 57.70, p < .001$. Comparisons of simple main effects of similarity within each modality condition indicated that with auditory presentation mean RT for HALV was significantly larger than that for LAHV ($p < .001$). However, with visual presentation the opposite relationship was obtained, so that the mean RT for HALV was significantly smaller than that for LAHV ($p < .01$).

TABLE 1

MEAN REACTION TIMES, STANDARD ERRORS, AND ERROR RATES (IN PERCENTAGES) AS A FUNCTION OF PRESENTATION MODALITY AND SIMILARITY

Modality	Similarity condition: HALV	Similarity condition: LAHV
Auditory		
$\bar{X}$	445	376
SE	20	21
Error	5	2
Visual		
$\bar{X}$	406	435
SE	22	25
Error	2	3

Note. Abbreviations: HALV = letter pairs high in auditory similarity and low in visual similarity; LAHV = letter pairs low in auditory similarity and high in visual similarity.

Discussion

The finding that similarity did have a reliable effect in both presentation conditions is evidence that the stimuli were coded and matched on modality-specific dimensions. Of prime importance is the fact that the RTs for the same contrasting HALV and LAHV pairs reliably varied as a function of presentation modality for the same *S*s in both modality conditions. Since similarity was the only other variable which was systematically manipulated, it seems reasonable to conclude that the differences in RT found for auditory and visual presentation are attributable to the similarity manipulations. The results suggest that the procedure could be expected to produce reliable differences in RT among the stimuli from which modality-specific coding could be inferred in further investigation. Hence, the same stimuli were used in a second experiment to study some additional questions related to auditory and visual coding.

EXPERIMENT II

As discussed previously, Tversky (1969) found that the same information was coded visually or verbally in a matching task depending upon the mode in which S expected S_2 to be presented. Thus, it appeared that the information from S_1 was recoded from one mode to the other when necessary and that the mode of S_1 had little effect on the matching task. However, Tversky used an ISI of 1 sec., which is at the upper limit of the estimated duration of the auditory (Massaro, 1970) and visual (Sperling, 1960) sensory stores. It is possible that at much shorter intervals, the residual sensory information from S_1 might have an effect on the dimension in which the 2 stimuli would be coded for matching. Hence, if the 2 stimuli were presented in different modalities, S might recode S_2 to the dimension in which S_1 was presented in order to match the two. To test this hypothesis in the present study, conditions were used in which both stimuli of a pair were either presented in the same or in different modalities (auditory or visual) at an ISI of approximately 150 msec. For comparison, the *same* conditions were presented at a longer ISI of 3 sec., at which sensory information should have decayed and S_1 would have to be maintained for matching through some type of rehearsal. The similarity manipulations described in Experiment I were used to assess the coding dimensions.

If sensory information can be utilized in matching, then RT should be faster at the short ISI than at the long ISI, where the 2 stimuli are presented in the same modality. However, in the conditions where the stimuli are presented in different modalities (mixed), the need to recode one of the stimuli to the modality

of the other for matching would lead to larger RTs at the short ISI than at the long ISI if the recoding takes longer than 150 msec. Likewise, the need for recoding would produce slower matches at the short ISI for the mixed-modality conditions than for the same-modality conditions.

An additional consideration regarding the ISI manipulation is related to the proposition that S_1 could be maintained for 3 sec. only by rehearsal. Such a need may cause *S* to use auditory coding even when both stimuli are always presented visually. Dainoff (1970) and Posner et al. (1969) did find a tendency toward auditory coding as ISI increased from 0–2 sec., although the mixture of "physical" and "name" match conditions may have influenced *S*s toward auditory coding, as indicated by Posner et al. (Experiment III). In the present study, a direct test of the relative effectiveness of auditory and visual coding was provided in the conditions where both letters of a pair were presented in the same modality at 2 ISI values and *S*s were instructed simply to indicate whether or not the 2 letters of a pair were the same.

Method

Subjects. The *S*s were 64 female students from the introductory psychology courses at the University of Iowa. They were tested individually in 2 50-min. sessions on successive days.

Design. The design was a 2 × 2 × 2 × 2 factorial combination of 2 between-*S*s variables and 2 within-*S*s variables. The between-*S*s variables were modality of S_1 (auditory or visual) and modality of S_2 (auditory or visual). Sixteen *S*s served in each of the 4 S_1 and S_2 modality combinations. The within-*S*s variables were similarity (HALV and LAHV) and ISI (short and long). The actual ISI values were chosen so that the short ISI (approximately 150 msec.) would be well within the estimated duration of the short-term sensory stores but the long ISI (3 sec.) would be well beyond it.

Materials. The materials were the same as those used in Experiment I.

Apparatus. The apparatus was the same as that used in Experiment I.

Procedure. In each experimental session, *S* received 5 blocks of trials, each consisting of a different random order of the 60 stimulus pairs plus 6 practice pairs at the beginning of each block. The 5 blocks of trials were first recorded for the long ISI condition with an ISI of 3 sec. and an intertrial interval (ITI) of 4 sec. For the short condition, a copy of the original tape was cut spliced between the stimuli of each pair result in ISI values with a mean of 160 msec. an standard deviation of 20 msec., as measured an oscilloscope. The 5 blocks of trials wit each session were presented in a different o to each *S*, and the order of the 2 ISI conditi was counterbalanced so that half of the *S*s rece the short ISI condition first and half received long ISI condition first. To aid *S* in discrimina between the ISI and the ITI, a light on the sti ulus panel was turned on during the ITI. T *S*s were randomly assigned to one of the 4 and S_2 modality conditions according to thei pearance at the laboratory, with the restrict that the total number of *S*s in each group .. differed by more than one.

The *S* rested the index finger of each hand one of the keys and was instructed to attempt make a correct response as quickly as possibl each trial, thus indicating whether or not S_1 and were the same letter. The *S* was also inform of the S_1 and S_2 modalities and that the modali would remain constant throughout the exp mental session. A 2-min. rest period was given tween each block of trials. The *same* response w always assigned to the dominant hand since th was no interest in directly comparing *same* a *different* responses.

Results

As in Experiment I, the first block of trials, the 6 practice trials in the remaining 4 blocks, and all error trials were excluded from the analysis. Error rates ranged 1%–7% across conditions. For each *S* a median RT was computed for each test pair, and then the 4 medians corresponding to each similarity condition (HALV or LAHV) were used for computing means. These data were then submitted to a 4-way analysis of variance.

The hypothesis that the modality of S_1 would affect coding at the short ISI, but not at the long ISI, predicts an interaction of all 4 variables. However, only the modality of S_2 affected coding as reflected by a reliable Similarity × S_2 Modality interaction, F $(1, 56) = 143.31$, $p < .001$. In the conditions where S_2 was auditory, the overall mean RT for HALV was 55 msec. larger than that for LAHV ($p < .001$); however, for the conditions where S_2 was visual, mean RT for HALV was 20 msec. smaller than that for LAHV ($p < .01$). Thus in general, the effects of

TABLE 2

MEAN REACTION TIMES, STANDARD ERRORS, AND ERROR RATES (IN PERCENTAGES) AS A FUNCTION OF S_1 MODALITY, S_2 MODALITY, INTERSTIMULUS INTERVAL (ISI), AND SIMILARITY

S_1 modality	Auditory S_2				Visual S_2			
	Short ISI		Long ISI		Short ISI		Long ISI	
	HALV	LAHV	HALV	LAHV	HALV	LAHV	HALV	LAHV
Auditory								
$\bar{X}$	454	411	522	460	369	402	431	438
SE	25	24	30	29	15	15	17	17
Error	3	1	6	1	1	1	1	2
Visual								
$\bar{X}$	426	370	457	395	446	465	471	490
SE	23	22	16	16	25	24	32	34
Error	5	1	7	1	1	1	1	1

Note. Abbreviations: HALV = letter pairs high in auditory similarity and low in visual similarity; LAHV = letter pairs low in auditory similarity and high in visual similarity.

similarity varied with the modality of S_2 regardless of the modality of S_1 or ISI. Mean RTs, *SE*s, and percentage error for the similarity conditions as a function of the other 3 independent variables are shown in Table 2.

One exception to the generalization that similarity effects varied as a function of S_2 modality occurred in the auditory S_1, visual S_2, long ISI condition, which led to a reliable Similarity × ISI × S_1 Modality interaction, F (1, 28) = 8.33, $p < .01$, in a separate analysis of the visual S_2 conditions. As shown in Table 2, the direction of the similarity difference was consistent with the other visual S_2 conditions but was only 7 msec., which failed to reach significance. One explanation for this result is that there were substantial similarity effects for each *S*, but the direction varied across *S*s. Individual *t* tests on the raw RTs for each *S*, however, failed to lend any support to this hypothesis. In general, the individual differences were rather small compared to the other visual S_2 conditions.

Earlier it was hypothesized that mixed-modality conditions might lead to slower matches than same-modality conditions at the short ISI, if the recoding of one stimulus to the modality of the other required more than 150 msec. To the contrary, however, a reliable S_1 Modality × S_2 Modality interaction, F (1, 56) = 6.03, $p < .025$, indicated that matches were faster in the mixed-modality conditions than where S_1 and S_2 were presented in the same modality. The *S*s responded 50 msec. faster in the visual–auditory conditions than in the auditory–auditory conditions and 58 msec. faster in the auditory–visual conditions than in the visual–visual conditions. Essentially the same pattern of results was also obtained from an analysis of *same* pairs.

Consistent with the hypothesis that sensory information can be used in matching, RTs at the short ISI were 40 msec. faster on the average than at the long ISI, F (1, 56) = 18.29, $p < .001$. The result was obtained for both the mixed- and same-modality conditions. Analysis of the *same* pairs again yielded an identical pattern of results.

Discussion

In using differences in RT to contrasting sets of letter pairs as a basis for making conclusions about coding, the assumption was made that the stimuli would be coded and matched on a common modality-specific dimension. The finding of reliable similarity effects in 7 of 8 conditions provides considerable support for that assumption. Since the similarity effects varied as a function of the S_2 modality in both the mixed- and same-modality conditions, it appears that the modality of S_1 had very little effect on the coding dimension, even at an ISI as short as 150 msec. Rather, *S*s still seemed to recode S_1 to the modality of S_2 when necessary, in

order to match the stimuli on the S_2 dimension. These results agree with those found by Tversky (1969), reviewed earlier, who used an ISI of 1 sec.

The agreement between Tversky's (1969) results and those of the present study is interesting in view of the procedural differences between the two other than the length of the ISI. First, Tversky randomly varied the mode of S_2 within an experimental session, whereas in the present study it remained constant. Second, Tversky used a considerably longer S_2 duration (4 sec.) than that of the present study (360 msec.). As reviewed earlier, Tversky's *S*s apparently matched the stimuli in the S_2 mode even when S_2 was presented in an unexpected (less frequent) mode, thus requiring a recoding of S_1 on those trials. It would seem that a recoding of the less frequent S_2 would be a more efficient strategy, but perhaps Tversky's use of the relatively long S_2 duration had an overriding influence toward matching in the S_2 mode. That is, a 4-sec. duration of S_2 would be more than enough time to recode S_1 as necessary and then sample the incoming sensory information from S_2 to complete the match. Thus, *S* would only be required to recode S_1 rather than to recode S_2 and retain S_1. Although a relatively short S_2 duration was used in the present study, matching in the S_2 dimension may have been due to keeping the S_2 modality constant during an experimental session. Thus the same results in the 2 studies may have occurred for different reasons.

One way in which a constant S_2 modality could have influenced the coding dimension in the present study is by providing *S* with a means to minimize the processing of S_2. This in turn would aid his following the instructions to respond as quickly as possible. Since the task was simply to decide if S_2 was the same stimulus as that previously presented, *S* could use information from S_1 to prime the perceptual mechanisms specific to a given modality for that same stimulus. In the same-modality conditions, the result of processing S_1 would merely have to be retained and used as a basis for testing the output from the same stage of processing S_2. The priming would be possible even when S_1 was presented in a different modality than S_2 but would require the additional processing necessary for recoding S_1. For example, an auditory S_1 would require a search of long-term memory for its visual equivalent, which could then be transferred to STM and used to prime the visual perceptual mechanisms for that particular letter. Facilitatory priming effects of advanced information have been found by Beller (1971) in a slightly different matching task for both same- and mixed-modality conditions.

Even though the modality of S_1 seems to have had very little effect on the matching dimension, there is some indication that residual sensory information from S_1 may have facilitated matching in the same-modality conditions. It is obvious th[illegible] in the long ISI conditions and the mixed-m[illegible]lity conditions sensory information from [illegible]ould not have played much of a role in [illegible] actual comparison process. In th[illegible] conditions the sensory information would [illegible]e decayed over the 3-sec. ISI and/or wou[illegible]ave only been a basis for recoding the stimu[illegible] to a different modality dimension. However, the finding that matches were faster at the short ISI than at the long ISI for the same-modality conditions is consistent with the notion that residual sensory information from S_1 could have been used to aid matching.

One result which ran counter to expectation was that matches in the mixed-modality conditions were faster than those in the same-modality conditions. The pattern of similarity effects indicates that S_1 was recoded to the modality of S_2 for matching in the mixed-modality conditions. Since the recoding would seem to require additional processing of S_1, the mixed-modality conditions were expected to be slower than the same-modality conditions rather than faster. Recent studies using memory match or memory search tasks (e.g., Burrows, 1972; Swanson, Johnson, & Briggs, 1972) have found RT in mixed-modality conditions to be either slower or no different from same-modality conditions. A tentative explanation for matches faster in the mixed- than in the same-modality conditions at the short ISI is that 150 msec. may not have been enough time for *S* to complete the necessary processing of S_1 when both stimuli were presented in the same modality, causing a delay in processing S_2. However, if it is assumed that the 2 modalities functioned as independent information channels that could be processed in parallel, there would be no delay in processing S_2 when it was presented in a different modality than was S_1. Evidence from studies using dichotic presentation of auditory stimuli (Norman, 1969; Peterson & Kroener, 1964)

provides some support for preliminary parallel processing of independent channels.

A final result of interest is related to the relative effectiveness of auditory and visual coding in STM. As stated earlier, the same-modality conditions provide a direct comparison of auditory and visual coding with all conditions equated except presentation modality. The finding that visual coding was used with visual presentation at both the long and short ISI seems to indicate that visual coding was as effective as auditory coding in the present study. These results are in agreement with those of other recent studies (Parks et al., 1972; Swanson et al., 1972) showing that *S*s can maintain visual information in STM beyond the duration of a sensory store, when there is a need to do so.

REFERENCES

Beller, H. K. Priming: Effects of advance information on matching. *Journal of Experimental Psychology*, 1971, **87**, 176–182.

Burrows, D. Modality effects in retrieval of information from short-term memory. *Perception & Psychophysics*, 1972, **11**, 365–371.

Clement, D. E., & Carpenter, J. S. Relative discriminability of visually-presented letter pairs using a same-different choice-reaction time task. *Psychonomic Science*, 1970, **20**, 363–365.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

Dainoff, M. J. Time course of visual and auditory encoding. *Journal of Experimental Psychology*, 1970, **86**, 214–224.

Dainoff, M., & Haber, R. N. Effect of acoustic confusability on levels of information processing. *Canadian Journal of Psychology*, 1970, **24**, 98–108.

Massaro, D. W. Perceptual auditory images. *Journal of Experimental Psychology*, 1970, **85**, 411–417.

McInish, J. R., & Tikofsky, R. S. Distinctive features and response latency: A pilot study. *Perception & Psychophysics*, 1969, **6**, 267–268.

Norman, D. A. Memory while shadowing. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 85–93.

Parks, T. E., Kroll, N. E. A., Salzberg, P. M., & Parkinson, S. R. Persistence of visual memory as indicated by decision time in a matching task. *Journal of Experimental Psychology*, 1972, **92**, 437–438.

Peterson, L. R., & Kroener, S. Dichotic stimulation and retention. *Journal of Experimental Psychology*, 1964, **68**, 125–130.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, **79**(1, Pt. 2).

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, **74**, 392–409.

Shulman, H. G. Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 499–508.

Sperling, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74**(11, Whole No. 498).

Swanson, J. M., Johnson, A. M., & Briggs, G. E. Recoding in a memory search task. *Journal of Experimental Psychology*, 1972, **93**, 1–9.

Tversky, B. Pictorial and verbal encoding in a short-term memory task. *Perception & Psychophysics*, 1969, **6**, 225–233.

Wickelgren, W. A. Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965, **70**, 102–108.

Wood, L. E. Visual and auditory coding in short-term memory. Unpublished doctoral dissertation, University of Iowa, 1971.

(Received April 9, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 114–120

EVENT OBSERVATION IN PROBABILITY LEARNING REVISITED

LOWELL M. SCHIPPER,[1] JACK A. THORPE, DAVID P. JONES,
AND GARY W. HEIMAN

Bowling Green State University

A 1968 study by Reber and Millward investigated the possible differential effect of observing as compared with predicting sequences in subsequent probability learning. The present study extends the range of probabilities and the numbers of observing trials. Specifically, 0, 10, 35, and 60 observing trials and πs of .60, .70, .80, and .90 were combined orthogonally. Additionally, different random orders were used for each observing and each predicting sequence for all *S*s. A rather firm basis is made for pointing up the lack of difference between observing and predicting procedures in some important dimensions.

Reber and Millward (1968) addressed themselves to the same basic problem investigated here: What is the effect of observing sequences of events as compared with predicting sequences of events on subsequent predictions in a traditional probability learning (PL) situation? Reber and Millward provide a more than adequate theoretical rationale for the problem and little more need be said here to justify the research.

Unfortunately, the Reber and Millward (1968) study is something less than definitive in providing an answer to the question as posed, for a number of reasons. Only a single probability for the most frequently occurring event, A_1, was used, $\pi = .80$. Only a relatively gross evaluation of the effect of observation sequences as compared with prediction sequences could be made since either 0, 100, 500, or 2,000 observation trials (OTs) were provided for comparison. Only a very few (4) sequences were used in predicting or testing and another 4 in observing or training. Additionally, but of minor importance, the observation treatments omitted the signal light which was present during test trials.

The purpose of the present study is to enlarge upon the Reber and Millward (1968) study using 4 different levels of π for A_1, .60, .70, .80, and .90, and 4 different amounts of OT, 0, 10, 35, and 60 trials. Since we had speculated that fewer OTs might be required with higher values of π for *S* to reach near-terminal levels of performance than with lower values of π, this arrangement makes it possible to examine this differential effect as a joint function of observation trials and π.

Additionally, our point of view is that of considering PL—unfettered by complex payoff contingencies or by stimuli with affective or with painful complications (as examples)—as a relatively "pure" form of cognitive learning or information processing, different from other types of learning such as conditioning or motor skills learning. That is, there is little involved in the input (stimulus) or output (response) of this learning situation that is of any consequence compared with the internal processing of the information conveyed by the simple occurrence of events. This point of view, obviously, sees PL as a rather unique or atypical form of learning requiring little or no perceptual learning, little or no response training, and little or no associating of special responses with novel stimuli. As a consequence, simple observations of events seem to provide all that is necessary for learning—or, more precisely, for changing proportions of A_1 predictions as a function of trials—to take place.

Method

Apparatus. Each *S* sat at a table positioned in front of a gray display panel on which were mounted 2 blue event lights (A_1 and A_2) 15.23 cm. apart at approximately eye level and a single red signal light 7.61 cm. above and centered between A_1 and A_2.

[1] Requests for reprints should be sent to Lowell M. Schipper, Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43403.

The response box on the table contained a center-off, double-throw lever switch, which was also centered between A_1 and A_2. Predictions of the occurrences of A_1 and A_2 were made by pushing the lever either to the left or right. The switch returned to center position when released by *S*.

Modified BRS-Foringer equipment controlled all timing, occurrences of A_1, A_2, and the signal light, and recording of all responses. Recording circuitry accepted as valid only those responses made during the particular 2-sec. interval when the signal light was on but before A_1 or A_2 appeared.

Design. Four levels of π for A_1 (.60, .70, .80, .90) were combined with 4 levels of OT (0, 10, 35, 60 trials) in a 4 × 4 factorial arrangement with 12 *S*s in each treatment. When appropriate, treatments will be identified as π/OT to indicate the particular combination of respective levels of each variable.

Three *E*s each ran 4 replicates of the experiment in random orders. Of the 4 *S*s for each *E* in each cell, 2 saw A_1 on the left and 2 on the right.

Procedure. For the 12 treatments with other than 0 OTs, each *S* simply saw either 10, 35, or 60 trials with either the left or right blue light coming on either 60%, 70%, 80%, or 90% of the time.[2] Of course, the complementary light appeared either 40%, 30%, 20%, or 10% of the time. During the OTs the signal light appeared for 1 sec. and was followed by an event light for 1.5 sec., which was followed by an intertrial interval of 1 sec. Following OTs each *S* received 100 test trials (TT) with the same π level that had been observed. Signal-light on time was increased to 2 sec. Other times remained the same. The 4 treatments receiving 0 OTs saw only 100 TTs at the appropriate π level.

All *S*s were run individually, one per session with *E* present. Essential parts of the instructions to *S*s were as follows.

With observation trials:

> Your job is to keep track of about how often each blue light comes on. That's all. Just keep track of how often each one comes on. It might seem as though the lights follow a *pattern* or *system*, but they *don't*. Just concern yourself with about how often each blue light comes on.

After observation trials were shown:

> Now this is important. For this series of trials, the *overall proportion* of lefts and rights will be exactly the same as the short series you just saw, although it will not be in the same order as the earlier series. Again, there will be no *pattern* or *system* although sometimes it might seem like there is. Your job is to be correct as often as possible. Try to be correct as many times as you can. Use what you learned about the proportions of lefts and rights in helping you predict as best you can.

Without observation trials (the principal portion of these instructions after the initial description of the procedure):

> Your job is to be correct as often as possible. Try to be correct as many times as you can. It might seem as though the lights follow a *pattern* or *system*, but they *don't*. Just concern yourself with getting as many correct predictions as you can.

Every one of the 144 sets of OTs was a different random arrangement, and every one of the 192 sets of TTs was also a different random arrangement.

Subjects. Female *S*s enrolled in introductory psychology courses at Bowling Green State University during the fall quarter of 1972 participated in this experiment as part of the psychology department's curricular requirements. Each *S* chose the experiment voluntarily and was permitted freedom to refuse to participate before and during her experimental session. None did.

Only a single sex was used in this sample of *S*s to prevent possible confounding of treatment with sex differences (Hanson & Schipper, 1971).

Results and Discussion

Figure 1 summarizes that portion of the data obtained during the 100 TTs for each S. The ordinate is the proportion of A_1 predictions; the successive abscissas are the first, second, and third thirds (33, 33, and 34 trials) of the 100 TTs; and the parameters for each subset of curves are the numbers of OTs. As expected, each subset fluctuates at or near the appropriate π level. The summary curves, one for each subset with its corresponding proportions of A_1 predictions attached, show no cross-overs. That is, the 3 points for .90 are all higher than the 3 points for .80, which are higher than .70, which are higher than .60. Regularity, vis-à-vis performance on TTs as a function of number of OTs, is *not* in evidence if one were to repeat the points of Figure 1 in subsets of 0 vs. 10 vs. 35 vs. 60 OTs with 4 curves in each family, one for each π value. Nor is there an orderly overall progression in response level as a function of increasing OTs from 0 through 60.

An analysis of variance of the data in Figure 1 using π and OTs as between-*S*s variables and blocks of trials (1–33, 34–66,

[2] Since 70% and 90% of 35 are 24.50 and 31.50, respectively, half of the *S*s in each of these treatments, half with each *E*, and half with left and right event lights received either 24 or 25 presentations of A_1 for .70 and either 31 or 32 presentations of A_1 for .90.

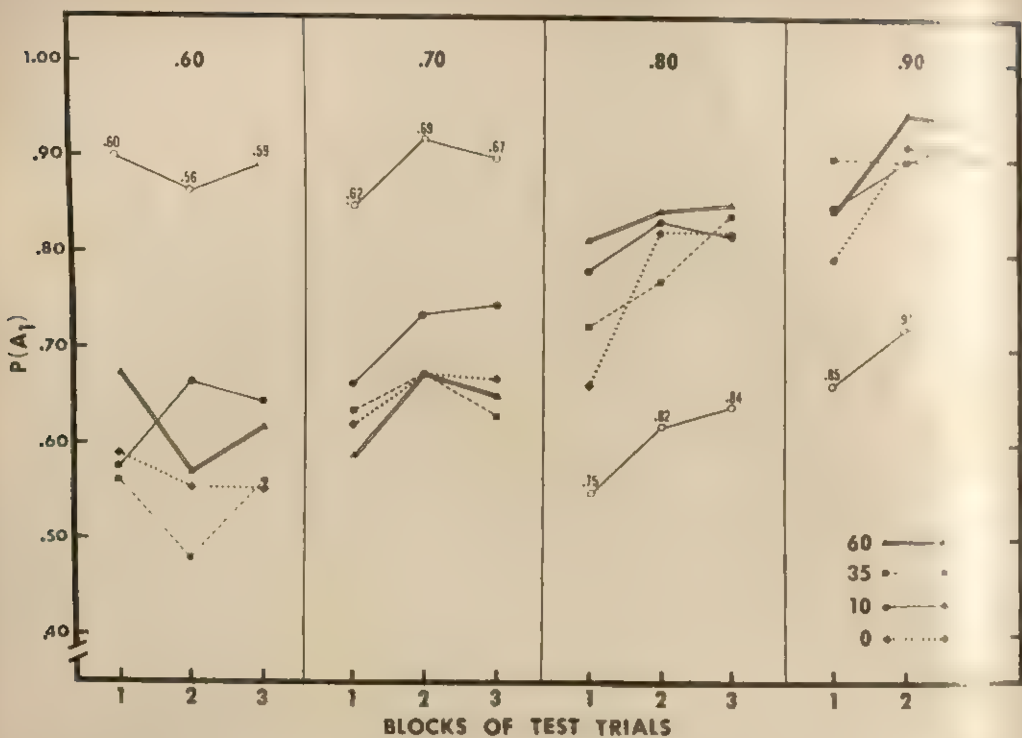


FIGURE 1. Proportion of A_1 predictions, $P(A_1)$, as functions of π (shown at the top of ea[illegible] [illegible]bset) and numbers of observation trials (shown in the lower right). (Curves with the values of th[illegible] [illegible]dinate attached reflect mean values for each subset. A_1 = most frequently occurring event.)

67–100) as a within-*S*s variable shows (*a*) differences in proportions of A_1 predictions as a function of π to be statistically reliable, $F(3, 176) = 92.39$, $p < .001$; (*b*) differences as a function of OTs to be reliable, $F(3, 176) = 3.02$, $p < .05$; and (*c*) blocks of trials, $F(2, 352) = 11.54$, $p < .001$, and the interaction of blocks and π, $F(6, 352) = 3.40$, $p < .005$, both reliable. The result of the first of these statistical tests is again expected and uninteresting. The third pair of tests gives statistical support to the interpretation of an increasing trend towards a higher response rate as a function of blocks with the caution that this trend is not the same for all π levels. The second statistical test, while showing that different numbers of OTs did bring about differencet in proportions of A_1 predictions, does nos expose the unexpected finding that the overall proportions are arranged 35 OTs (.713) < 0 OTs (.714) < 60 OTs (.750) < 10 OTs (.762) in increasing order, which shows no *systematic* (and little overall) variation attributable to numerosity of OTs.

Perhaps the most informative analysis of OTs vs. TTs is the use of staggered blocks of trials as originally suggested by Reber and Millward (1968). Two such analyses are presented here.

At each of the levels of π, Trials 67–100 for the 0 OT treatments, Trials 57–90 for the 10 OT treatments, Trials 32–65 for the 35 OT treatments, and Trials 7–40 for the 60 OT treatments were compared. This procedure essentially uses the data for the last block of 34 TTs in the 0 OT treatments as a referent and makes comparisons among (*a*) 66 TTs, (*b*) 10 OTs plus 56 TTs, (*c*) 35 OTs plus 31 TTs, and (*d*) 60 OTs plus 6 TTs (see top half of Figure 2). If observing and predicting have the same effect, then response curves for these different blocks for all levels of π should be

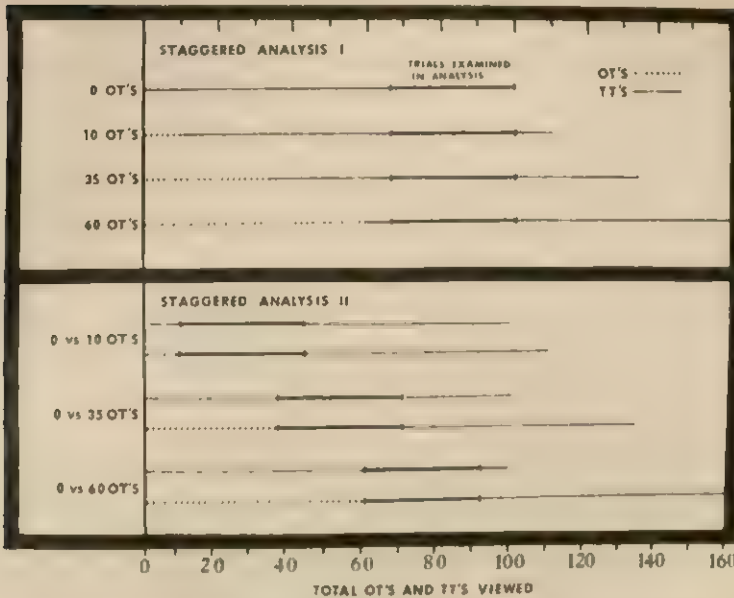


Figure 2. Test trials used for comparisons in first and second staggered analyses. (OT = observation trial; TT = test trial.)

horizontal. A summary of the results is shown in Figure 3.

... dinate in Figure 3 is again propor- ... predictions and the abscissa is ... r of observation trials. These ... re rather flat for $\pi = .80$ and ... ut rather nonflat for $\pi = .60$ and π ...

Treating these data as entries in a simple 4×4 factorial arrangement (with $n = 12$ in each cell), the differences in the 4 levels of the curves as a function of π yield F (3, 176) = 58.20, $p < .001$, and once again a predictable and expected finding. Differences between numbers of OTs yield F (3, 176) = 3.48, $p < .05$, which indicates

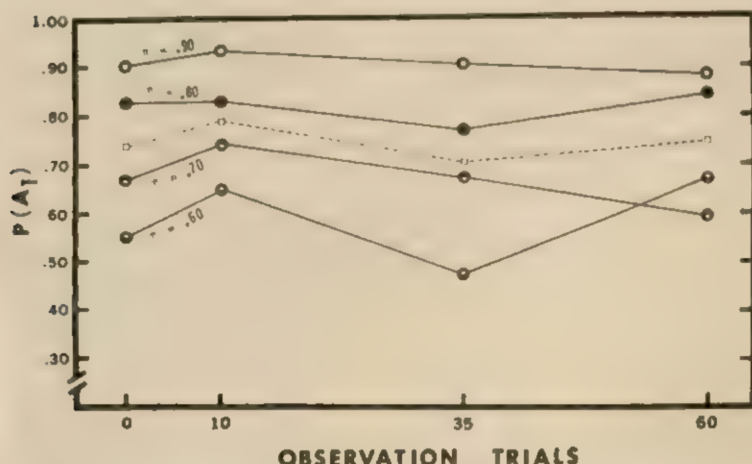


Figure 3. Staggered analysis showing comparisons among various combinations of observation and prediction trials. (The dashed curve is fitted to the 4 means at each level of observation trial. $P(A_1)$ = proportion of A_1 predictions; A_1 = most frequently occurring event.)

TABLE 1

PROPORTION OF A_1 RESPONSES AS A FUNCTION OF PRECEDING OBSERVATION AND TEST TRIALS AND AS A FUNCTION OF π

π	10		35		60	
	Observation	Test	Observation	Test	Observation	Test
.90	.853	.910	.900	.913	.850	.885
.80	.788	.738	.724	.825	.818	.838
.70	.662	.681	.630	.678	.584	.666
.60	.575	.562	.560	.542	.673	.571
M	.719	.723	.703	.739	.731	.740

Note. Entries for the Observation cells at the respective π values are mean proportion of A_1 (most frequently occurring event) responses for Test Trials 1–33. Entries for the corresponding Test cells are mean proportion of A_1 responses for Test Trials 11–43 when 10 observation trials were used for comparison; for Test Trials 36–68 when 35 observation trials were used for comparison; for Test Trials 61 93 when 60 observation trials were used for comparison. For example, the entry in the .90/10 Observation cell is the mean for *S*s who had received 10 observation trials prior to the measured 33 test trials. The entry in the .90/10 Test cell is the mean for *S*s who had received 10 test trials prior to the 33 measured test trials. Each entry for the Observation cells is based on 12 different *S*s. Each entry for Test cells .90/10, .90/35, and .90/60 is based on only those 12 *S*s who received the .90/0 treatment. Similarly, the .80/10, .80/35, and .80/60 Test cells are based on the 12 *S*s who made up the .80/0 treatment, and so forth.

an overall deviation from flatness when data are averaged over π levels and a single point is plotted for each of the OTs. The interaction of π and OTs, F (9, 176) = 2.05, $p < .05$, is also reliable and represents, of course, the lack of parallelness of the 4 curves shown in Figure 3.

A second staggered analysis, somewhat more complicated than the first, is illustrated in the bottom half of Figure 2. Essentially what this analysis does is provide a look at TT performance shifted farther and farther "down the line" in the TT sequence as a function of the number of OTs substituted for TTs. Table 1 summarizes this analysis, and a complete description of the nature of each entry is given with the table.

To be noted in Table 1 is the similarity between the means for 10 OT and 10 TT treatments, .719 and .723; for 35 OT and 35 TT, .703 and .739; and for 60 OT and 60 TT, .731 and .740, with 95% confidence intervals for the differences between each of these 3 pairs of means equal to ±.054, ±.048, and ±.049, respectively. Differences appear to be truly minute and suggest only a slight (if any) difference in a positive direction toward the programmed mean π of .75.

Statistical tests of differences between pairs of columns in Table 1, 10 OT vs. 10 TT, 35 OT vs. 35 TT, and 60 OT vs. 60 TT, with the 4 values of π as a second variable, uniformly show statistically reliable differences for π (again, as would be expected) and give F ratios of less than unity for 10 OT vs. 10 TT and 60 OT vs. 60 TT, and F (1, 88) = 2.18, $.10 < p < .25$, for 35 OT vs. 35 TT. All 3 of the interactions of π and type of trials give F ratios with confidence levels exceeding .05, and only one of these, that for 60 trials, is close to .05.

Still another analysis, that of trends, can be carried out to answer the question of whether or not the collective set of curves in Figure 2 is flat, linear with slope other than zero, quadratic, or cubic in form. According to a hypothesis of no difference between OTs and TTs, flatness is expected.

Using a procedure outlined in Saito (1965) for unequally spaced values on the abscissa, only the cubic form of the polynomial provides an F ratio which "accounts for" a statistically significant portion of the variance attributable to numbers of OTs, F (1, 176) = 8.71, $p < .005$. This cubic expression reflects the 2 inflections which are apparent when plotting only the 4 means as shown by the dashed curve in Figure 3.

A final look at what occurred during TTs is given in Table 2. This table shows the distributions of *S*s along all dimensions of the study during the successive thirds as well as during the complete set of 100 TTs. Deserving of special attention are the distributions of variances for 0, 10, 35, and 60 OT treatments at successive stages of the experiment. These OTs represent the principal variable in the experiment, and lack of homogeneity here could invalidate tests of significance among their means.

Cochran's C statistic with $k = 4$, 47 *df*, and $\alpha = .05$ for each appropriate set of 4 variance estimates shows that there is no basis for rejecting a hypothesis of homogeneous variability at any of the 3 stages of TTs, nor among the overall distributions.

Overall, we feel there is little difference between types of trials in terms of subsequent performance in the traditional PL

TABLE 2

FREQUENCY DISTRIBUTIONS OF ALL *S*s

% A1 responses	0					10					35					60				
	60	70	80	90	All	60	70	80	90	All	60	70	80	90	All	60	70	80	90	All
Test trials 1–33[a]																				
91–100	0	1	0	2	3	1	1	5	6	13	0	0	1	8	9	2	1	3	5	11
81–90	1	0	1	4	6	0	0	0	2	2	0	1	3	4	8	0	1	4	5	10
71–80	3	3	4	3	13	2	3	3	3	11	2	4	3	0	9	1	4	2	0	7
61–70	3	1	4	2	10	0	4	3	1	8	2	3	4	0	9	5	2	3	2	12
51–60	0	3	2	1	6	6	3	1	0	10	5	1	0	0	6	4	3	0	0	7
41–50	4	4	0	0	8	0	1	0	0	1	0	2	1	0	3	0	0	0	0	0
31–40	1	0	1	0	2	2	0	0	0	2	3	1	0	0	4	0	1	0	0	1
21–30	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
11–20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0–10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Test trials 34–66[b]																				
91–100	0	0	4	7	11	1	0	5	7	13	0	1	2	6	9	0	2	4	9	15
81–90	0	3	3	4	10	1	3	3	4	11	0	2	6	4	12	0	0	4	3	7
71–80	0	2	3	1	6	1	4	2	0	7	1	2	0	1	4	2	2	3	0	7
61–70	5	3	2	0	10	5	4	2	1	12	2	4	1	1	8	4	6	1	0	11
51–60	3	2	0	0	5	3	1	0	0	4	4	1	3	0	8	2	0	0	0	2
41–50	3	2	0	0	5	1	0	0	0	1	1	1	0	0	2	1	2	0	0	3
31–40	1	0	0	0	1	0	0	0	0	0	2	0	0	0	2	3	0	0	0	3
21–30	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0
11–20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0–10	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
Test trials 67–100[c]																				
91–100	0	1	4	7	12	0	1	4	10	15	0	0	4	6	10	0	0	3	11	14
81–90	0	2	1	4	7	1	2	2	1	6	0	3	4	6	13	1	2	4	0	7
71–80	2	2	6	1	11	3	4	4	1	12	4	3	3	0	10	2	3	5	1	11
61–70	3	3	1	0	7	3	5	1	0	9	2	2	1	0	5	4	3	0	0	7
51–60	3	1	0	0	4	4	0	1	0	5	1	1	0	0	2	2	1	0	0	3
41–50	2	3	0	0	5	1	0	0	0	1	2	1	0	0	3	2	2	0	0	4
31–40	2	0	0	0	2	0	0	0	0	0	2	1	0	0	3	1	1	0	0	2
21–30	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
11–20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0–10	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
Test trials 1–100[d]																				
91–100	0	0	1	2	3	0	0	1	5	6	0	0	0	4	4	0	0	2	7	9
81–90	0	0	4	9	13	1	3	6	5	15	0	2	7	7	16	1	2	5	5	13
71–80	0	3	4	1	8	1	3	4	1	9	0	2	2	1	5	1	0	5	0	6
61–70	4	6	3	0	13	5	5	1	1	12	4	4	2	0	10	5	6	0	0	11
51–60	6	1	0	0	7	4	1	0	0	5	2	3	1	0	6	3	2	0	0	5
41–50	2	2	0	0	4	1	0	0	0	1	5	0	0	0	5	2	2	0	0	4
31–40	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
21–30	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
11–20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0–10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Note. The entries in the respective 60, 70, 80, and 90 columns show the distributions of those 12 *S*s giving A1 (most frequently occurring event) responses who had received 0, 10, 35, and 60 observation trials. The 4 All columns show the distributions of all 48 *S*s with 0, 10, 35, and 60 observation trials.

[a] Variances of the All distributions in the 0, 10, 35, and 60 observation trials conditions are σ^2 = 253.78, 338.89, 332.64, and 228.48, respectively.

[b] Variances of the All distributions in the 0, 10, 35, and 60 observation trials conditions are σ^2 = 300.20, 199.83, 466.62, and 335.38, respectively.

[c] Variances of the All distributions in the 0, 10, 35, and 60 observation trials conditions are σ^2 = 320.79, 208.16, 414.40, and 312.33, respectively.

[d] Variances of the All distributions in the 0, 10, 35, and 60 observation trials conditions are σ^2 = 195.14, 166.49, 295.66, and 241.49, respectively.

situation. We agree with Estes' (1972) interpretation of the Reber and Millward (1968) experiment saying approximately the same thing. Some differences may be a function of the particular π levels used, with these differences probably observable near $\pi = .50$ but not near extreme values.

REFERENCES

Estes, W. K. Research and theory on the learning of probabilities. *Journal of the American Statistical Association*, 1972, 67, 81–102.

Gatto, J. Unequal intervals and une[illegible] end analyses. *Psychological Bulletin*, [illegible] 63, 125–127.

Hansson, B. L., & Schipper, L. M. N[illegible], in probability learning and decision [illegible] with multiple equivalent predictors. [illegible] *Science*, 1971, 22, 199–201.

Reber, A. S., & Millward, R. B. Event [illegible]rvation in probability learning. *Journal o[illegible] perimental Psychology*, 1968, 77, 317–327.

(Received April 12, 1973)

Manuscripts Accepted for Publication in the

Journal of Experimental Psychology

Rotary Acceleration of a Subject Inhibits Choice Reaction Time to Motion in Peripheral Vision. J[illegible] M Borkenhagen (Department of Psychology, University of Southern California, University P[illegible] Los Angeles, California 90007).

Relative Effectiveness of Rhymes and Synonyms as Retrieval Cues. Douglas Lee Nelson (Depart[illegible]t of Psychology, University of South Florida, Tampa, Florida 33620) and David Heber Brooks.

A Semantic Interpretation of Encoding Specificity. Lynne M. Reder (Department of Psychology, [illegible]man Performance Center, University of Michigan, Ann Arbor, Michigan 48104), John R. Anders[illegible] and Robert A. Bjork.

Shock-Associated Words in a Nonattended Message: A Test for Momentary Awareness. R. S. C[illegible]rteen (Department of Psychology, University of British Columbia, Vancouver, British Columbia, C[illegible]nada) and D. Dunn.

Interactions and Range Effects in Experiments on Pairs of Stresses: Mild Heat and Low Frequency Noise. E. C. Poulton (MRC, Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, England) and R. S. Edwards.

Preconditioning the Effects of Shock-Correlated Reinforcement. John G. Carlson (Department of Psychology, University of Hawaii, 2430 Campus Road, Honolulu, Hawaii 96822).

Recognition Time for Words in Short-Term, Long-Term, or Both Memory Stores. Richard C. Atkinson (Department of Psychology, Stanford University, Stanford, California 94305) and Richard C. Mohs.

The Effect of the Range of Test Stimulus on Stimulus Generalization in Human Subjects. David R. Thomas (Department of Psychology, Muenzinger Building, University of Colorado, Boulder, Colorado 80302), Gary Hansen, Arthur Tomie, and Doris H. Thomas.

Storage Coding Tradeoff in Short-Term Store. Francis S. Bellezza (Department of Psychology, Porter Hall, Ohio University, Athens, Ohio 45701) and Richard J. Walker.

Short-Term Memory for Sounds and Words. Edward J. Rowe (Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada), Ronald P. Philipchalk, and Leslie J. Cake.

Judgment of Recency for Pictures and Words. Gary L. Lassen (Department of Psychology, University of Arizona, Tucson, Arizona 85721), Terry C. Daniel, and Neil R. Bartlett.

Stimulus and Response Frequency and Sequential Effects in Memory Scanning Reaction Times. John Theios (Department of Psychology, University of Wisconsin, Psychology Building, Charter at Johnson, Madison, Wisconsin 53706) and Dennis G. Walter.

Variables Affecting the Intermanual Transfer and Decay of Prism Adaptation. Chong Sook Choe (Department of Psychology, University of Kansas, Lawrence, Kansas 66044) and Robert B. Welch.

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 121–124

RETRIEVAL OF SUPERORDINATES AND SUBORDINATES [1]

ELIZABETH F. LOFTUS [2]
University of Washington

MARTIN BOLTON
New School for Social Research

Semantic memory retrieval was investigated in 2 experiments. In the first experiment, *S*s were shown a category name and asked to respond with a word belonging to the category (for one block of trials) or a class to which the category name belonged (for another block of trials). Subjects produced a category member faster than they produced a superordinate. For example, they produced an instance of the category CAR faster than they produced a superordinate such as *vehicle*. The time taken to retrieve a superordinate was strongly related to the category's hierarchical position, while the time taken to retrieve an instance was not so related. In the second experiment, *S*s produced free associates to categories differing in hierarchical level. More subordinates were given for all but the lowest level categories. The data argue against the notion that the superset is the most accessible property of a category or concept.

The present article deals with relations of class inclusion, relations which have been of interest at least since Aristotle's time. Knowledge of these inclusion relations is knowledge that we have learned some time ago and know very well; it forms a part of our semantic memory. Whenever we make a subordinate or superordinate response, as when we assert that (*a*) an example of a dog is a collie, or (*b*) the class to which collies belong is "dogs," we manifest our knowledge of inclusion relations.

Class inclusion relations have become central in many current theories of memory. In the network model of Rumelhart, Lindsay, and Norman (1972), for example, a statement such as "A car is a vehicle" is represented as 2 nodes, one corresponding to "car" and the other to "vehicle," and a relation between them, the "isa" relation. One question that arises is what directional effects in information retrieval exist as a consequence of these inclusion relations? For example, given the stimulus word CAR, is it easier to retrieve a superordinate response, such as *vehicle*, *object*, or *thing*, or it is easier to retrieve an instance, such as *Ford*, *Buick*, or *Ferrari?*

Miller (1969) claims that "a superordinate response will be much stronger (more frequent and probably faster) than a subordinate response [p. 231]." He points to results from word association tests to support his claim. For example, the results of a reclassification of the data of Kent and Rosanoff (1910) indicated that 7.6% of all the responses given were superordinate and only 1.6% were subordinate (Woodrow & Lowell, 1916). From this evidence, Miller concluded that the subordinate–superordinate direction of association is stronger than the superordinate–subordinate direction. Collins and Quillian (1972) seem to be in agreement with Miller, as they say "in many cases the superset is the most accessible property of a concept. . . . In contrast, the instances of a concept are not easily accessible properties in general [p. 320]."

Rather than trust "indirect" data from word association tests to settle the matter of directionality, an experiment was performed in which *S*s were shown a word on each trial and were asked to name an instance (for one block of trials) or a superordinate (for another block of trials). The *S*s were timed while they produced their responses.

Experiment I

Method

Subjects. The *S*s were 22 students at the New School for Social Research. Each *S* took part in one experimental session that lasted 30 min.

[1] The authors are grateful to Sif Wiksten for her assistance in data analysis.

[2] Requests for reprints should be sent to Elizabeth F. Loftus, Department of Psychology, University of Washington, Seattle, Washington 98195.

TABLE 1

Median Reaction Time (in sec.) Taken to Retrieve a Superordinate and a Subordinate for Twenty Sets of Nested Categories

Level 1 category			Level 2 category		
Name	Superordinate	Subordinate	Name	Superordinate	Subordinate
CLOTHING	2.61	1.54	FOOTWEAR	2.10	[illegible]
SCHOOL	2.25	1.67	COLLEGE	1.93	[illegible]
ACTIVITY	4.61	1.71	CRIME	3.80	[illegible]
SPORT	1.86	1.42	WATERSPORT	3.00	[illegible]
FOOD	2.74	1.99	VEGETABLE	1.95	[illegible]
VEHICLE	3.58	1.30	CAR	1.92	[illegible]
TYPE OF BUILDING	2.65	1.89	RELIGIOUS BUILDING	2.82	[illegible]
CITY	2.58	1.19	U.S. CITY	3.08	[illegible]
TITLE	3.37	1.92	MILITARY TITLE	3.79	[illegible]
PLANT	2.54	2.14	TREE	2.01	[illegible]
POLITICIAN	3.03	1.88	PRESIDENT	2.05	[illegible]
SCIENCE	3.51	1.59	NATURAL SCIENCE	3.61	[illegible]
MINERAL	4.06	2.79	METAL	2.97	[illegible]
BEVERAGE	2.23	1.50	SOFT DRINK	2.30	[illegible]
STONE[a]	3.45	3.49	GEM	1.66	[illegible]
TYPE OF MONEY	2.61	1.59	COIN[a]	1.44	[illegible]
GEOMETRIC SHAPE	2.62	1.54	TRIANGLE[a]	2.04	[illegible]
ANIMAL	2.31	1.30	DOG[a]	1.40	[illegible]
WORD	2.28	1.73	NOUN[a]	2.02	[illegible]
FURNITURE	5.78	1.35	CHAIR[a]	1.44	[illegible]
M	3.08	1.72		2.29	[illegible]

[a] Categories for which production of superordinates was faster than production of instances.

Materials. Twenty pairs of nested category names were constructed. For each pair of categories, the superordinate category (Level 1) included everything that belonged in the subordinate category (Level 2). The nested categories used are presented in Table 1. The category names were printed on 5 × 8 in. index cards, with one category per card. Each *S* received 2 random permutations of the 40 cards. Half of the *S*s named superordinates on the first 40 trials, and subordinates on the last 40 trials. The reverse arrangement held for the remaining *S*s.

Procedure. Each *S* was told that we were conducting a study on how memory works, that he would see items consisting of single words, and that he was to respond with a word that was a superordinate (subordinate) of the stimulus word. He was given examples and told to respond as quickly as possible, but to avoid errors.

The *S* sat in front of a screen in which there was a window covered by half-silvered glass. The index card containing the stimulus was placed in a dark enclosure behind the mirror and was presented by illuminating the enclosure. A microphone was placed in front of *S* and he responded by speaking into it.

A trial consisted of the following events. As a card with the item printed in large type was placed in the darkened enclosure behind the half-silvered mirror, *E* said "Ready," and pressed a button that illuminated the stimulus. The *S*'s verbal response activated a voice key that stopped the clock and terminated the trial. If *S* did not respond within 10 sec., the trial was terminated. A warm-up period of 10 trials, using stimuli different from those used for the experimental trials, preceded each block of experimental trials.

Results

Only correct responses (96%) to the 20 pairs of nested categories are included in the following analyses. For each of the nested pairs, one category is more inclusive (higher in the semantic hierarchy) than the other. For example, VEHICLE includes more instances than CAR; in some sense, CAR is lower in the hierarchy than VEHICLE. We designated the higher category "Level 1" and the lower category "Level 2," and we computed a Level 1 and Level 2 median reaction time (RT) for each *S* for each of the 2 types of responses (subordinate and superordinate).

A 2 × 2 repeated measures analysis of variance indicated that the speed of producing a subordinate was significantly faster than the speed of producing a superordinate, F (1, 21) = 46.46, $p < .01$. All 22 *S*s showed this effect. Furthermore,

there was a significant effect of hierarchical position of a category on the time taken to respond, F (1, 21) = 9.86, $p < .01$. The interaction was also significant, F (1, 21) = 12.94, $p < .01$, indicating that the mean time taken to produce a superordinate was influenced by category level, while the time to produce a subordinate was not so influenced.

Median RTs were computed for subordinate and superordinate responses to each of the 40 categories and are presented in Table 1. For 32 stimulus words, *S*s produced a subordinate faster than he produced a superordinate; for 6 stimuli, the reverse result occurred.

Discussion

The finding that subordinates are produced more quickly than superordinates is inconsistent with Miller's (1969) and Collins and Quillian's (1972) intuitions that a superordinate is more accessible than a subordinate. On the other hand, Woodworth and Wells (1911) found that *S* produces superordinates faster than he produces subordinates, which is, of course, opposite to our own result. To reconcile their data with the present results, we must examine the Woodworth and Wells procedure more closely. In their experiment, 13 *S*s named superordinates of 20 stimulus words (for example, *S*s produced a superordinate for OAK and for CABBAGE, etc.) and they named subordinates of 20 *different* words (for example, TREE and VEGETABLE). A comparison of the average time taken for the 2 types of responses indicated a shorter RT for naming superordinates.

Although the present procedure is sounder than that of Woodworth and Wells (1911) in that we use the same stimulus words for the production of both superordinates and subordinates, the discrepancy between the 2 studies must be explained. A closer examination of the Woodworth and Wells stimulus words indicated that at least 12 of the words in the "name a superordinate" test were themselves instances of the words in the "name a subordinate" test. For example, *S*s were required to name a superordinate of OAK, CABBAGE, and PENNY, but to name subordinates of TREE, VEGETABLE, and COIN. Not only were the words for which *S*s produced superordinates lower in any assumed hierarchical structure than the words for the "subordinate test," but these words were lower than the words we used in Experiment I. Thus, the words for which Woodworth and Wells' *S*s produced superordinates would constitute, in our terminology, a "Level 3." These observations suggested the possibility that the discrepancy between the Woodworth and Wells study and the present study was due to item differences (specifically in hierarchical level) between the 2 studies. To test this possibility, Experiment II was designed.

Experiment II

Method

Subjects. The *S*s were 22 college students, none of whom had participated in Experiment I.

Materials. The 20 pairs of nested categories shown in Table 1 were used. In addition, a group of Level 3 categories was created by selecting one subordinate for each of the Level 2 categories. So, for example, for the pair FOOD–VEGETABLE, we selected the stimulus LETTUCE. Each *S* was presented a random permutation of the 60 unique categories.

Procedure. The *S*s were told that we were conducting an experiment on free association. They were told that they would see a series of 60 stimulus words, and after seeing each word, they were to write down the first word that it made them think of.

Results

A simple count was made of all response words which were superordinate and those which were subordinate. In case of any doubt, the simple question "Does one of these words belong to the class of things named by the other word?" would usually solve the problem. This technique has been used successfully by Peters (1952). In Figure 1, the mean numbers of these 2 types of responses are presented separately for the 3 levels of stimulus words. A 2 (response type) $\times$ 3 (hierarchical level) $\times$ 22 (*S*s) analysis of variance indicated that more subordinates than superordinates were given, F (1, 21) = 43.90, $p < .01$. Also significant were hierarchical level F (2, 42) = 32.94, $p < .01$, and the Level $\times$ Response Type interaction, F (2, 42) = 170.37, $p < .01$. A 2 (response type) $\times$ 3 (levels) $\times$ 20 (category triplet) analysis of variance gave identical results, namely, significant response type, level, and interaction effects.

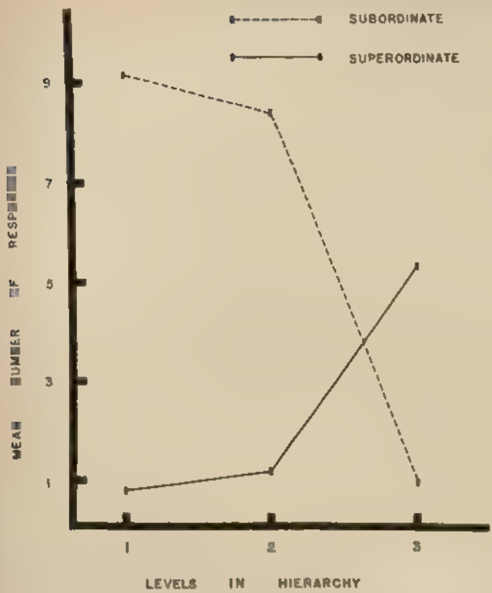


FIGURE 1. Mean number of subordinate and superordinate responses given in Experiment II.

General Discussion

Whether or not the superordinate or subordinate is the stronger depends on the level of the stimulus word in the semantic hierarchy. At Levels 1 and 2, more subordinates are given as free associates (Experiment II) and subordinate responses are given faster (Experiment I). At the lowest level, Level 3, more superordinates are given as free associations, and presumably superordinate responses would be given faster. At this level, *S*s appear to have few (if any) subordinate responses available. The reader can verify this fact by attempting to produce some instances of CABBAGE or PENNY. Subordinate responses are unlikely as free associates, but, if forced, *S* might reply with *red cabbage* or *copper penny*. Responses are highly idiosyncratic and correct single-word responses rarely occur.

The finding in Experiment I that the mean time taken to produce a subordinate when given a Level 1 category was not different from the time taken to produce a subordinate to a Level 2 category replicates an earlier study of Loftus, Freedman, and Loftus (1970), which showed that the RT taken to retrieve a member of a given category was not significantly different from the RT taken to retrieve a member of a superset of that category. Hierarchical position of a category does, however, affect the speed of producing superordinates. The RT taken to produce a superordinate of a Level 2 category such as CAR was faster than the RT taken to produce a superordinate of a Level 1 category such as VEHICLE. Thus, the more specific a word is, the faster *S*s can produce a superordinate for that word. In fact, an examination of the 6 categories listed in Table 1 for which production of superordinates was faster than production of instances reveals that 5 of the categories are the lower member of their respective nested pair. Thus, not only do some Level 2 (or specific) categories have a readily accessible superordinate, but they do not have as accessible an instance. There are so few of these cases (and in all 6, there is less than .8-sec. difference in the RT to produce a subordinate vs. a superordinate), that we find it difficult to agree with any overall statement that "the instances of a concept are not easily accessible [Collins & Quillian, 1972, p. 320]." In fact, the results of both experiments point out the need for recognizing that statements such as "the subordinate–superordinate direction of association is stronger" (cf. Miller, 1969) are, at best, only sometimes true

References

COLLINS, A. M., & QUILLIAN, M. R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

KENT, G. H., & ROSANOFF, A. J. A study of association in insanity. *American Journal of Insanity,* 1910, **67**. 37–96, 317–390.

LOFTUS, E. F., FREEDMAN, J. L., & LOFTUS, G. R. Retrieval of words from subordinate and superordinate categories in semantic hierarchies. *Psychonomic Science,* 1970, **21**, 235–236.

MILLER, G. A. The organization of lexical memory: Are word associations sufficient. In G. A. Talland & N. C. Waugh (Eds.), *The pathology of memory.* New York: Academic Press, 1969.

PETERS, H. N. Supraordinality of associations and maladjustment. *Journal of Psychology,* 1952, **33**, 217–225.

RUMELHART, D. E., LINDSAY, P. H., & NORMAN, D. A. A process model of long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

WOODROW, H., & LOWELL, F. Children's association frequency tables. *Psychological Monographs,* 1916, **22**(5, Whole No. 97).

WOODWORTH, R. S., & WELLS, F. L. Association tests. *Psychological Monographs,* 1911, **13**(5, Whole No. 57).

(Received April 13, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 125-132

PUNISHMENT OF APPETITIVELY REINFORCED INSTRUMENTAL BEHAVIOR:

FACTORS AFFECTING RESPONSE PERSISTENCE

DENNIS G. DYCK[1]
Pepperdine University

ROGER L. MELLGREN
University of Oklahoma

JACK R. NATION
University of Oklahoma

In the first of 2 runway investigations, 2 groups of rats received equal amounts of punishment and nonreinforcement during acquisition training. Group PR received transitions from punishment to reinforcement and Group NR received transitions from nonreinforcement to reinforcement. A control group received unpunished continuous-reinforcement training. Following training, half of the *Ss* in each group received punished extinction, and the other half received unpunished extinction. The results indicated that Group PR was superior to Group NR in punished extinction, and the reverse was true in unpunished extinction. The control group exhibited the poorest performance regardless of the response-decrement procedure employed. In the second experiment, Groups NR and PR received unpunished, punished, and unpunished extinction following acquisition. Generally, the results were consistent with those obtained in the first experiment. These results were interpreted in terms of Capaldi's 1967 sequential theory of instrumental learning.

It is a well-established finding that partial punishment increases response persistence to continuous punishment over and above that produced by partial reinforcement (e.g., Banks, 1966a, 1966b; Brown & Wagner, 1964). Response persistence is increased more readily when punishment (P) is administered on nonreinforced (N) trials rather than reinforced (R) trials (Fallon, 1968, 1969), although persistence effects (relative to unpunished controls) have been observed in situations where punishment has been presented simultaneously with reinforcement (i.e., Banks, 1966b; Brown & Wagner, 1964).

Persistence effects due to punishment have traditionally been interpreted via an extension (Wagner, 1966) of Amsel's (1962, 1967) analysis of frustrative nonreward. However, several recent attempts have been made to extend Capaldi's (1967) sequential theory to account for the effects of punishment upon instrumental performance (Campbell, Crumbaugh, Marshall, & Sparling, 1972; Campbell, Crumbaugh, Massey, & Reed, 1972; Campbell, Wroten, & Cleveland, in press; Capaldi & Levy, 1972).

According to a sequential analysis, the mechanism of increased persistence depends upon the sequence of different goal-box events and not on the "number" of such events per se (Capaldi & Kassover, 1970). In a partial-reinforcement situation, resistance to extinction is increased when the memory of nonreinforcement (S^N) is conditioned to the instrumental response (R_I). Sequence is an important variable due to the fact that S^N is conditioned to R_I only on reinforced trials that have been preceded by nonreinforced trials.

Assuming that punishment results in a memory (S^P) like S^N, it may be conditioned to R_I on reinforced trials that follow punished trials (P–R transitions). It follows that differential trial sequencing should produce differential response persistence. This hypothesis has been supported in 2 separate investigations (Campbell et al., in press; Capaldi & Levy, 1972). In both studies, it was found that response persistence to punished extinction was increased

[1] Requests for reprints should be sent to Dennis G. Dyck, who is now at the Department of Psychology, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.

by partial-punishment schedules containing P–R transitions relative to schedules containing only R–P transitions. Thus, preliminary evidence suggests that P–R transitions increase response persistence in a manner similar to N–R transitions. In addition, Capaldi and Levy found that groups which received punishment prior to large reward performed better over a series of punished trials than a group which received punishment prior to small reward. This finding indicates that the sequence of punishment–reinforcement events, not the ordinal position of punishment in the daily series of trials, determined performance.

The present experiments were designed to further explore the effects of sequential manipulations using both punishment and nonreinforcement. Accordingly, in the first of 2 experiments, 2 groups of *S*s experienced partial punishment–partial reinforcement training, so that one group (Group PR) received P–R transitions and the other group (Group NR) received N–R transitions. A third group received continuous reinforcement (Group RR). It should be noted that both experimental groups received the same number of nonreinforcement–reinforcement–punishment events; only the trial sequence of these events was varied. Following training, the 3 groups were split in half so that half of the *S*s in each group received punished extinction and the other half received unpunished extinction. The unpunished-extinction and punished-extinction procedures will be referred to as *response-decrement procedures.* It was hypothesized that differential group performance would be a function of the interaction between the type of training sequence and the kind of response-decrement procedure used in testing. Specifically, it was assumed that the stimulus complex in punished extinction would be dominated by punishment stimuli (S^P) and therefore *S*s trained with theoretical S^P–R_I associations (Group PR) should be superior to *S*s trained with theoretical S^N–R_I associations (Group NR). Conversely, the opposite prediction was made for *S*s receiving unpunished extinction. That is, Group NR was expected to show greater persistence relative to Group PR in unpu[illegible] extinction (i.e., since unpunished [illegible]tion would occasion nonreinforceme[illegible]ory stimuli). Of course, Group R[illegible] expected to show the least pers[illegible] regardless of the response-decr[illegible] procedure employed.

Experiment 1

Method

Subjects. The *S*s were 60 experime[illegible]y naive male albino rats of the Sprague-Daw[illegible] strain, purchased from the Holtzman Co. [illegible] were housed individually and were approx[illegible]ely 70 days old at the start of the experiment[illegible] The *S*s were randomly assigned to 1 of 3 ex[illegible]mental groups ($n = 20$ group).

Apparatus. The apparatus consisted [illegible] single straight-alley runway manufactured by [illegible]Hunter. The alley, which was 150 × 15 × 9 cm., [illegible]as constructed of clear Plexiglas with a grid flo[illegible] It was divided into a 30-cm. start section, a [illegible]m. run section, and a 30-cm. goal section, all sec[illegible]s being separated by guillotine doors. A teaspoon [illegible]ounted in the middle of the far end of the goal [illegible] served as a food cup. The *S*'s progress in the [illegible]ey was measured by 3 .01-sec. Standard Electric [illegible] Corp. timers. The first timer, which measured [illegible] time, was started by a microswitch at the start-box door and stopped by a photocell located 11 cm. into the alley. The second timer, which measured run time, was started by the first photocell and stopped by a second photocell located 11 cm. in front of the goal box. The third timer, which measured goal time, was started by the second photocell and stopped by a third photocell located 9 cm. inside the goal box. Start, run, and goal speeds were obtained by converting the start, run, and goal times to reciprocals. Total speeds were obtained by summing the start, run, and goal times and reciprocating this measure. Punishment consisted of a .3-ma. scrambled shock which was gradually increased to .5 ma. for .5 sec., administered in the goal section of the alley by a Model 700 Grason-Stadler shock generator. The shock was administered manually by pushing a button on the apparatus control panel.

Procedure. For 3 days after arrival in the laboratory, *S*s were allowed free access to ad-lib food and water. The *S*s were then placed on a 12-gm. daily food-deprivation schedule, with water continuously available. The deprivation schedule was established 7 days before the start of the experiment. During the 3 days before the start of the experiment, each *S* was handled individually for 3–5 min. daily. Two days of pretraining preceded the experiment proper. On the first day all *S*s received 2 reinforced trials, and on the second day they received 4 continuously reinforced trials. The food cup in the goal box was baited on reinforced trials throughout the experiment with 2 1-cm.

Purina Hog Starter pellets (approximately 90 mg. each).

Two groups received 66% reinforcement, 17% nonreinforcement, and 17% punishment, and an additional group received 100% reinforcement. In Group NR, all nonreinforced trials were followed by reinforced trials, and punished trials were never followed by reinforced trials. In Group PR, all punished trials were followed by reinforcement, and nonreinforced trials were never followed by reinforcement. Group RR received 100% reinforcement throughout training. Punishment was administered immediately after *S* broke the final photobeam. The goal box was not baited on punished trials, and *S*s were removed from the goal box following punishment. The intensity of punishment was gradually increased from .3 ma. to .5 ma. in .1-ma. increments during every fifth day of training. On nonreinforced trials, *S*s were confined in the goal box for 15 sec.

Acquisition training lasted 15 days with 4 trials/day. The rats were run in squads of 6, with a 15-sec. intertrial interval. In order to prevent *S*s from forming a pattern discrimination, 3 different sequences were used during acquisition, each being used a total of 5 times. The 3 sequences of N, R, and P trials for Group NR were NRRR, RRRP, and NRRP. For Group PR, the sequences were RRRN, PRRR, and PRRN. Following acquisition, half of the *S*s in each group received punished extinction and the other half received unpunished extinction. Testing consisted of 20 trials in 5 days, 4 trials/day.

Results

Acquisition. Asymptotic performance was evaluated by analyzing the mean daily running speeds over the last 5 days of acquisition (20 trials). A 3 (Group) × 2 (Response Decrement) × 5 (Days) repeated-measures analysis of variance indicated nondifferential group performance as a function of training in start, F (2, 54) = 1.76, $p > .05$; in run, F (2, 54) = .38, $p > .05$; and in total, F (2, 54) = 2.54, $p > .05$. The main effect of group did account for a significant portion of the variance in the goal measure, F (2, 54) = 9.64, $p < .01$. Post hoc comparisons (all comparisons reported in this and the subsequent experiment used the Scheffé correction procedure for post hoc comparisons) indicated that Group RR was superior to both Group PR and Group NR ($p < .01$). Unlike the acquisition data reported by Capaldi and Levy (1972), Groups PR and NR did not differ from each other. The response decrement variable (extinction vs. punished extinction), which was a dummy variable for acquisition, was nonsignificant in all measures (all Fs < 1). The main effect of days was significant in start, F (4, 216) = 6.10, $p < .01$; in goal, F (4, 216) = 26.02, $p < .01$; and in total, F (4, 216) = 8.68, $p < .01$; but not in the run measure, F (4, 216) = 2.16, $p > .05$. None of the interactions in the analysis were significant. As the analyses indicated, very little response suppression was observed in Groups NR and PR relative to the continuously reinforced controls. It should be mentioned that at no time did any *S*s in Groups NR or PR evidence patterning. Mean speeds across the last 5 days of acquisition are shown as Point A in Figure 1.

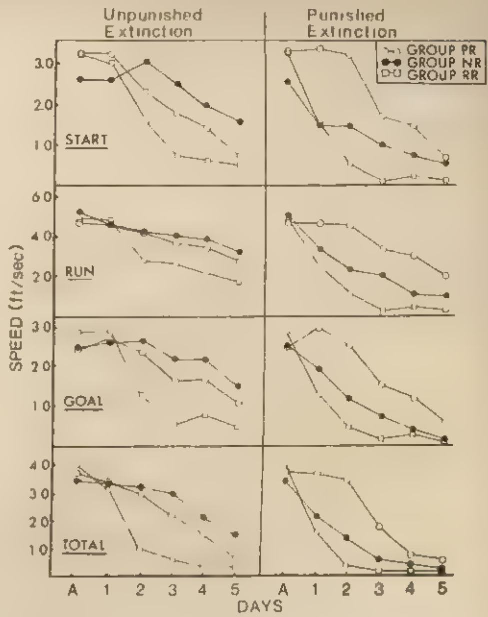


FIGURE 1. Mean running speeds for the last day of acquisition (A) and 5 days of punished and unpunished extinction for all runway measures.

Response decrement. A 3 (Groups) × 2 (Response Decrement) × 5 (Days) analysis on the mean daily speeds indicated a significant main effect for groups in start, F (2, 54) = 19.47, $p < .01$; in run, F (2, 54) = 46.21, $p < .01$; in goal, F (2, 54) = 31.87, $p < .01$; and in total, F (2, 54) = 50.59, $p < .01$. Post hoc comparisons

showed that PR groups were superior to NR groups, and both were superior to RR groups ($p < .01$ in all cases). The response decrement main effect (punished extinction vs. unpunished extinction) accounted for a significant portion of the variance in all sections of the alley, the smallest F being 22.15 in start ($df = 1, 54$ in all cases). This finding reflects the fact that punished extinction produced greater response decrements than unpunished extinction. The most important finding was the significant Group × Response Decrement interaction in start, $F(2, 54) = 8.20$, $p < .01$; in run, $F(2, 54) = 11.12$, $p < .01$; in goal, $F(2, 54) = 10.38$, $p < .01$; and in total, $F(2, 54) = 14.40$, $p < .01$. Post hoc comparisons indicated that Group PR was superior to Group NR in punished extinction in all measures. On the other hand, Group NR was superior to Group PR in unpunished extinction in all measures ($p < .01$ in all cases) except run, where the 2 groups did not differ from each other. Of course, Group RR was inferior to all other groups, regardless of the response-decrement procedure employed. It is notable that Group PR performed as well in punished extinction as in unpunished extinction in some sections of the alley. In start and total measures, performance to the 2 response-decrement procedures was nondifferential; however, performance in the run and goal sections was better in unpunished extinction than in punished extinction ($p < .05$ in both cases). Both Group NR and Group RR were superior in unpunished extinction relative to punished extinction. Of course, the days main effect was highly significant in all sections. The Fs for start, run, goal, and total were 55.16, 82.49, 91.29, and 124.34, respectively ($df = 4, 216$ in all cases). The greatest decrement was noted in the goal section and the least in start, with run showing intermediate suppression. Finally, the Groups × Days interaction was significant in all sections, with the smallest F being in run, $F(8, 216) = 3.33$, $p < .01$. This interaction is indicative that the groups extinguished at different rates. These effects for each of the alley segments across days are shown in Figure 1.

Discussion

The acquisition data repo[illegible]d here differ somewhat from previous res[illegible]n (Campbell et al., in press; Capaldi & Le[illegible] 1972) in that suppression, which was very [illegible]ght, was observed only in the goal section [illegible] in both of the studies mentioned, suppressio[illegible] was substantial during training, and th[illegible] [illegible] suppression was dependent on th[illegible] [illegible] sequential transitions experienced. It is [illegible]ely that the lack of suppression in this exp[illegible]ment was due to the fact that punishment w[illegible] administered on a sparse schedule (i.e., on[illegible] 17% of the trials were punished). In ad[illegible] [illegible] punishment was given on every thi[illegible] [illegible] training. It is notable that suppression [illegible] observed occurred in the goal section. [illegible], that segment of the instrumental res[illegible] chain occurring in the closest tempor[illegible] and spatial proximity to punishment wa[illegible] suppressed whereas earlier segments of the [illegible]ponse were not. Evidence from a number [illegible] experiments (Campbell & Meyer, 1971; Cap[illegible]di & Levy, 1972; Capaldi & Ziff, 1969) p[illegible] support for the notion that mildly ave[illegible] stimuli in the goal box produce differe[illegible] between groups in the goal section, whereas more intense aversive stimuli produce group differences in the earlier segments of the runway. Presumably, differences in the start and run sections of the alley between the punished and unpunished groups would have been observed if punishment intensity and/or density had been increased in the present experiment.

The response-decrement results obtained in the present experiment reflect the powerful influence of "sequence" in determining response persistence. More importantly, response persistence was shown to be a function of the interaction between the type of sequences used in training and the kind of response-decrement procedure used in testing. Consistent with this prediction, it was found that Group PR was superior to Group NR in punished extinction, while the reverse was true in unpunished extinction. Thus, P–R transitions increase response persistence relative to N–R transitions when testing is done with punishment-related stimuli. Conversely, N–R transitions provide for greater persistence than P–R transitions when testing occurs with nonreinforcement-related stimuli. This interaction is predicted by sequential theory, following the assumption that persistence to response-decrement testing is increased by the similarity between memory stimuli conditioned to the

instrumental response during training and the stimuli presented during testing (i.e., stimulus-generalization decrement). Thus, both Group PR and Group NR performed better when they were tested with stimuli that had been conditioned to the instrumental response during training than when they were tested with stimuli not directly conditioned. It is noteworthy that, in a similar experiment, Campbell et al. (in press) found that *S*s trained with both N–R and P–R transitions were as persistent in unpunished extinction as *S*s trained only with N–R transitions. A comparison of the present results with those reported by Campbell et al. illustrates how subtle variations in "sequence" can dramatically alter performance.

It is notable that Group PR was markedly superior to Group NR in punished extinction, but the superiority of Group NR over Group PR was less pronounced in unpunished extinction. In fact, the run measure did not reflect differential group performance in unpunished extinction. This finding may be interpreted by considering that punishment was administered on nonreinforced trials. For Group PR, subsequent reinforced trials may have resulted in the conditioning of a compound memory stimulus consisting of both nonreinforcement and punishment elements (S^N plus S^P) to R_I. Thus, Group PR was provided with some degree of persistence to nonreinforcement-related stimuli as well as to punishment-related stimuli. This analysis also provides an explanation for the group main effect finding which showed Group PR to be more persistent than Group NR when the data were collapsed across the response-decrement variable.

Experiment II

Although the data from Experiment I clearly reflect the powerful influence of sequential manipulations in determining response persistence, it was felt that an additional experiment would provide more information on the interaction between training sequence and response-decrement procedure.

Thus, 2 groups were given training similar to that experienced by Groups NR and PR in Experiment I. Following training, all *S*s experienced a 3-phase response-decrement test, in which a period of punished extinction was administered between 2 periods of unpunished extinction.

Due to the brevity of the initial phase, it was hypothesized that performance would be nondifferential. In the second phase (punished extinction) it was hypothesized that, due to the reinstatement of punishment-related cues, Group PR would be superior to Group NR. Finally, in the third phase of the test it was hypothesized that Group NR might be superior to Group PR. The third phase (unpunished extinction) was included to determine whether or not the reinstatement of S^N would produce increased performance in Group NR.

Method

Subjects. The *S*s were 20 experimentally naive male albino rats of the Sprague-Dawley strain, purchased from the Holtzman Co. They were housed individually and were approximately 70 days old at the start of the experiment. The *S*s were randomly assigned to 1 of 2 experimental groups ($n = 10$/group).

Apparatus. The apparatus was identical to that utilized in Experiment I.

Procedure. Acquisition training was carried out in the same manner as in Experiment I, with the following exceptions: (*a*) training was extended to 112 trials, 4 trials/day for 28 days, and (*b*) several additional sequences were included to prevent *S*s from forming a pattern discrimination. The additional sequences for Group NR were RRNR and RNRP. For Group PR, the additional sequences were RRPR and RPRN. The percentage of nonreinforcement–reinforcement–punishment events was exactly the same as in Experiment I. Furthermore, it should be noted that on any given day, all *S*s experienced the same number of each kind of goal-box event.

Response-decrement testing consisted of 3 phases: unpunished extinction, punished extinction, and unpunished extinction. Testing consisted of 24 trials, 4 trials/day for 6 days. Each of the 3 phases lasted for 2 days.

Results

Acquisition. The mean daily running speeds during the last 5 days of acquisition were analyzed by means of a 2 (Group) × 5 (Days) repeated measures analysis of variance. The differences between Groups NR and PR were negligible at the end of acquisition training in start, goal, and total measures (*F*s < 1), although there was a

nonsignificant trend for Group PR to be superior to Group NR in the run section, F (1, 18) = 3.58, $p > .05$. The days main effect was significant, indicating a general increase in speeds, F (4, 72) = 5.09, $p < .01$; in run, F (4, 72) = 10.19, $p < .01$; in goal, F (4, 72) = 4.83, $p < .01$; in total, but not in start, F (4, 72) = 1.88, $p > .05$. Finally, the interaction did not approach significance in any of the runway sections, the largest F being in total, F (4, 72) = 1.69, $p > .05$.

Response decrement: Phase 1 (unpunished extinction). A 2 (Group) × 2 (Days) repeated measures analysis of variance, performed on the mean daily running speeds, indicated a nonsignificant main effect for groups in all runway measures: Fs = .01, .69, 1.06, and .55 in start, run, goal, and total sections, respectively (df = 1, 18 in all cases). The days main effect was significant at the .01 level or better in all measures, the smallest F being 9.23 in start (df = 1, 18). The Groups × Days interaction was negligible in all runway measures (Fs < 1 in most sections). Thus, the first phase of the response-decrement test produced a nondifferential decline in response speed (i.e., both groups were running more slowly than in aquisition, but they were not different from each other).

Response decrement: Phase 2 (punished extinction). A 2 (Group) × 2 (Days) analysis indicated a significant main effect for groups in start, F (1, 18) = 5.46, $p < .05$; in run, F (1, 18) = 11.02, $p < .01$; in goal, F (1, 18) = 5.93, $p < .05$; and in total, F (1, 18) = 8.06, $p < .01$. The days main effect was significant in all segments of the runway with the smallest F being in start, F (1, 18) = 39.03, $p < .01$. The Group × Days interaction did not account for a significant portion of the variance in any of the runway sections. The major finding was that Group PR was statistically superior to Group NR in punished extinction, consistent with the finding in Experiment I.

Response decrement: Phase 3 (unpunished extinction). A 2 (Groups) × 2 (Days) analysis on the mean daily running speeds (8 trials) revealed a significant group main effect in run, F (1, 18) = 4.72, $p < .05$; in goal, F (1, 18) = 7.19, $p < .05$; and in total, F (1, 18) = 4.89, $p < .05$; but not in start, F (1, 18) = 3.32, $p > .05$. Group PR maintained its superiority over Group NR despite the fact that no reinforcement stimuli were reinstated in the goal box without punishment. The days main effect accounted for a significant portion of the variance only in the start measure, F (1, 18) = 12.15, $p < .01$. This indicates that *S*s had reached asymptotic extinction in all runway segments except for the start measure, in which some decline was still observed. The Group × Days interaction was negligible in all runway sections (most Fs < 1).

Phase 2 trial analysis. In order to examine Phase II performance more closely, a 2 (Group) × 4 (Trials) analysis was performed on the running speeds for the 4 punished trials on each of the 2 days of punished extinction. The analysis on the first day of punished extinction showed that the group main effect was significant in all runway measures: in start, F (1, 18) = 5.67, $p < .05$; in run, F (1, 18) = 13.35, $p < .01$; in goal, F (1, 18) = 8.36, $p < .01$; and in total, F (1, 18) = 10.19, $p < .01$. The Group × Trials interaction also accounted for a significant portion of the variance in start, F (3, 54) = 2.92, $p < .05$; in run, F (3, 54) = 3.13, $p < .05$; and in total, F (3, 54) = 5.87, $p < .01$; but not in goal F (3, 54) = 1.29, $p > .05$. Post hoc comparisons indicated that the performance of the 2 groups on the first punished trial was nondifferential. However, Group PR was superior to Group NR on the second trial (all ps < .01) and all succeeding trials. Thus the superiority of Group PR over Group NR was evidenced after the administration of only one punished trial. Of course, the trials main effect was significant in all runway measures. On the second day of punished extinction, a pattern of results similar to the first day was observed except that performance was slower and therefore the magnitude of the group differences was smaller. The group main effect was reliable in start, F (1, 18) = 4.56, $p < .05$, and in run, F (1, 18) = 7.81, $p < .05$; but not in goal, F (1, 18) =

3.49, $p > .05$, or total, $F(1, 18) = 4.19$, $p > .05$. The Group $\times$ Trials interaction was statistically significant in start, $F(3, 54) = 4.04$, $p < .05$; in run, $F(3, 54) = 7.14$, $p < .01$; and in total, $F(3, 54) = 2.91$, $p < .05$; but not in goal, $F(3, 54) = 2.56$, $p > .05$. Appropriate post hoc comparisons (on start, run, and total data) indicated that group performance was nondifferential on the first punished trial: however, by the second trial, Group PR was superior to Group NR ($p < .01$ in all cases except goal). The trial-by-trial performance of the groups during the 2 days of punished extinction is shown in Figure 2. An inspection of the graphs on the second day of punished extinction indicates that Group PR actually ran faster on the second trial than on the first. This effect was statistically reliable in both the start and run sections ($p < .05$ and $p < .01$, respectively). On the other hand, it can be seen that Group NR ran reliably slower on the second trial of both punished extinction days ($p < .01$ in all cases). A related finding was that the groups performed nondifferentially on the first trial of each day, thereby indicating that spontaneous recovery was not differently affected by the treatment variables.

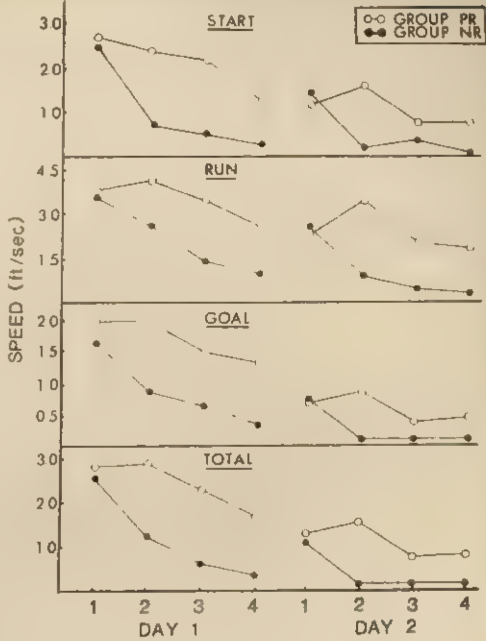


FIGURE 2. Mean running speeds for the 2 days of punished extinction for all runway measures.

Discussion

The response-decrement findings in Experiment II are generally consistent with the results of Experiment I. Following a block of unpunished extinction trials (Phase 1) in which performance was nondifferential, the sudden introduction of punishment produced dramatic between-groups performance differences (Phase 2). These differences were nonexistent on the first punished trial, but appeared in robust form on following trials. The sudden superiority of Group PR over Group NR in punished extinction indicates that stimulus control (i.e., the capacity of S^P to occasion R_I) of punished reactions is maintained through a block of unpunished extinction trials.

In contrast, the data obtained in the final phase of unpunished extinction did not conform to prediction. It was hypothesized that the reinstatement of S^N would result in increased performance for Group NR and decreased performance for Group PR. This hypothesized interaction did not occur; rather, both groups seemed to reach asymptotic extinction performance with Group PR still running significantly faster than Group NR. Several possibilities may be considered to account for the Phase 3 data. First, it is possible that proactive interference occasioned by Phase 2 punishment resulted in the disruption of the S^N–R_I association for Group NR. This is a plausible assumption, but it is not immediately clear why unpunished extinction (Phase 1) did not proactively interfere with Phase 2 performance for Group PR. A second possibility to account for the poor performance of Group NR in Phase 3 is that nonreinforcement memory stimuli are simply not as distinctive as punishment memory stimuli, and therefore stimulus control established with S^N dissipates or is disrupted more readily than stimulus control established with S^P. Finally, it is possible that Phase 3 performance merely reflects *S*'s ability to recover from the effects of response-contingent punishment. Group PR had reinforcement associated with the aftereffect of punishment during training, whereas Group NR did not; consequently S^P may have been psychologically more intense for the latter *Ss*. If S^P was a more "intense"

event for Group NR relative to Group PR, poorer recovery for the latter group would be expected. There is evidence (Boe & Church, 1967) that recovery from punishment does not occur (when punishment is delivered at the beginning of extinction) at a wide range of punishment intensities, with the extent of suppression being inversely related to punishment intensity. In summary, it is possible that a combination of the factors mentioned is responsible for the poor performance exhibited by Group NR in Phase 3.

General Discussion

These experiments suggest that the trial sequencing of partial-punishment partial-reinforcement training controls performance during response-decrement testing. Furthermore, the level of performance depends not only on the trial sequence of nonreinforcement–reinforcement–punishment events in training, but also upon the kind of response-decrement procedure that is used during testing. These effects were demonstrated using both a between-*Ss* and a within-*Ss* response-decrement procedure.

These data confirm and extend the results of Campbell et al. (in press) and Capaldi and Levy (1972) and lend further support to the notion that sequence is a crucial determinant of response persistence. It appears that P–R transitions function in a manner analogous to N–R transitions. However, the theoretical similarity between nonreinforcement and punishment could be extended even further if effects analogous to those obtained in partial-reinforcement–extinction situations were demonstrated using punishment.

References

Amsel, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review*, 1962, 69, 306–328.

Amsel, A. Partial reinforcement effects on vigor and persistence. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 1. New York: Academic Press, 1967.

Banks, R. K. Persistence to continuous punishment and nonreward following training with intermittent punishment and nonreward. *Psychonomic Science*, 1966, 5, 105–106. (a)

Banks, R. K. Persistence to c[illegible] punish[illegible]ment following intermittent p[illegible] training. *Journal of Experimental Ps[illegible]* 1966, 71, 373–377. (b)

Boe, E. E., & Church, R. M[illegible] [illegible]nt effects of punishment during extinct[illegible] [illegible] *Comparative and Physiologica[illegible]* [illegible], 1967, 63, 486–492.

Brown, R. T., & Wagner, A[illegible] [illegible]stance to punishment and extinction [illegible] training with shock or nonreinforce[illegible] *Journal of Experimental Psychology*, 1964 [illegible] 507.

Campbell, E. M., & Meyer, [illegible] Effects of daily reward sequence on sim[illegible] and successive negative contrast in [illegible] *Journal of Comparative and Physiologica[illegible]* [illegible], 1971, 74, 434–440.

Campbell, P. E., Crumbaugh, C[illegible] [illegible] R. L., & Sparling, C. L. A p[illegible] [illegible] tion based on intermittent p[illegible] [illegible] *Psychological Reports*, 1972, 30, 1005[illegible]

Campbell, P. E., Crumbaugh, C[illegible] [illegible] & Reed, E. Facilitation of an [illegible] crimination by shock on N tria[illegible] [illegible] *Reports*, 1972, 30, 333–334.

Campbell, P. E., Wroten, J. D., & [illegible] T. Response persistence follow[illegible] [illegible] punishment and partial reinforc[illegible] [illegible] trial sequence. *Learning and M[illegible]* [illegible] in press.

Capaldi, E. J. A sequential hyp[illegible] [illegible] of instru[illegible]mental learning. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning* [illegible] Vol. 1. New York: Academic Pr[illegible] 1967.

Capaldi, E. J., & Kassover, K. S[illegible]ce, number of nonrewards, anticipation, and i[illegible]trial interval in extinction. *Journal of Exp[illegible]ntal Psychology*, 1970, 84, 470–476.

Capaldi, E. J., & Levy, K. W. Stimulus control of punished reactions: Sequence of punishment trials and magnitude of reinforcement trials. *Learning and Motivation*, 1972, 3, 1–19.

Capaldi, E. J., & Ziff, O. R. Schedule of partial reward and the negative contrast effect. *Journal of Comparative and Physiological Psychology*, 1969, 68, 593–596.

Fallon, D. Resistance to extinction following learning with punishment of reinforced and nonreinforced licking. *Journal of Experimental Psychology*, 1968, 76, 550–557.

Fallon, D. Resistance to extinction following partial punishment of reinforced and/or nonreinforced responses during learning. *Journal of Experimental Psychology*, 1969, 79, 183–185.

Wagner, A. R. Frustration and punishment. In R. N. Haber (Ed.), *Current research in motivation.* New York: Holt, Rinehart & Winston, 1966.

(Received April 23, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 133 141

VERBAL CONTEXT SHIFTS AND FREE RECALL [1]

ALAN S. BROWN AND BENTON J. UNDERWOOD [2]

Northwestern University

The number of different stimuli (1, 2, 3, 4, 9, 18, or 36) associated with 36 response terms was manipulated as a means of simulating variation in contextual shifts in free recall. Recall was inversely related and pair recognition directly related to number of different stimuli. When response terms with the same stimulus terms were blocked within a list, performance generally decreased within blocks. When the pairs were unblocked, evidence for unlearning of the early pairs was found. The interpretation of unlearning was supported by the further finding that spontaneous recovery of the initial items occurred over a 20-min. interval. Insofar as the present procedures simulated context shifts realistically, it must be concluded that a complex set of phenomena emerge as context shifts in free recall.

Investigators representing a wide range of theoretical persuasions have found it convenient, if not necessary, to invoke contextual stimuli to account for at least some of the responses produced in single-trial free recall. Contextual stimuli are usually said to be represented by the characteristics of the physical environment or by *S*'s emotional and cognitive activity, the latter including the implicit verbal responses which may be produced by the words presented for learning. Direct evidence on the role of contextual stimuli in memory has sometimes been produced (e.g., Turvey & Egan, 1969), but there is nearly a complete lack of systematic studies on the topic for free-recall learning. The original conception of the first study to be reported here viewed the conditions as being ones which would provide systematic evidence on the role of context. As will be seen, they may also be viewed as dealing with other related phenomena.

Consider a situation in which *S* is presented a paired-associate list for one trial and is then asked to recall only the response terms. The formal stimulus terms may be conceived of as contextual stimuli, each serving as a unique context for its response term. Like contextual stimuli, the stimulus terms do not have to be overtly produced. If they serve a role in the recall of the response terms, their systematic status is essentially the same as that attributed to context cues except that, in this case, they are under experimental control. If *S* can recall (to himself) a stimulus term, and if an association has been established between it and the response term, the response term will be produced. Of course, the critical question is sidestepped, for an appeal to some other context would be necessary to explain how the stimulus term was recalled. Only if the contextual stimuli are always present (as would be true for the physical properties of the room) could this problem be avoided. Any notion of a nonstatic context—a shifting context or a series of successive contexts—requires that in some way that context be reinstated (remembered) if it is to serve as an effective cue for recall.

Although the present experiment does not resolve the basic issue, it proceeds on the assumption that it is quite possible to vary the likelihood with which context stimuli will be recalled. In the first experiment, *S*s were presented a paired-associate list of 36 pairs. There were 36 different stimulus terms at one extreme, and in successive conditions there were fewer and fewer different stimulus terms, each being paired with more than one response term, until at the other extreme there was a single stimulus term paired with 36 different response terms. The pairs with common

[1] The authors acknowledge with gratitude the critical appraisal of this report by John J. Shaughnessy.

[2] Requests for reprints should be sent to Benton J. Underwood, Department of Psychology, Northwestern University, Evanston, Illinois 60201.

stimulus terms always occurred in blocks. It might be assumed that contextual shifts occur on the study trial of a long free-recall list of single words. The present manipulations were intended to simulate such shifts; each change in a stimulus term was likened to a context shift. Furthermore, it can be seen that as the number of context shifts decreases, the frequency with which each stimulus term (context) is presented increases. Thereby, the probability that the stimulus term will be recalled increases.

If the formal stimulus terms are to simulate contextual cues in free recall, 2 learning conditions must be met. First, *S* must know that he will be expected to recall the response terms without the stimulus terms (context) present. Second, at the same time, he must know that he will also be held responsible for associating the response terms with the context. Therefore, in addition to being instructed that response-term recall would be required, *S* was told that his knowledge of the pairings would be tested by a recognition procedure.

If the procedures as described are a reasonable representation of context variation, it follows that context changes within a single free-recall study trial could bring into play a number of phenomena, listed below.

1. When more than one response term is paired with a single stimulus term, an A-B, A-C paradigm is set up within the list. This could produce either or both of 2 outcomes. First, the earlier learned pairs within a block could produce negative transfer in attempts to learn later pairs. Second, and working in the opposite direction, attempts to learn later pairs in a block could produce unlearning of earlier pairs. Furthermore, the magnitude of the effects ought to vary as a function of the number of different response terms paired with a common stimulus term.

2. One could view each block of pairs with the same stimulus term as a list. Thus, when the same stimulus term occurred with 12 different response terms there would be 3 different lists. Since retroactive inhibition has been shown for free-recall lists, a compar[illegible] phenomenon may emerge within the l[illegible]

3. An increase in the [illegible] of different stimuli within the list m[illegible] reduce the interference effects resulti[illegible]om the A-B, A-C paradigm but, at the [illegible] time, increase the memory load [illegible] of different words) for *S*.

These potential com[illegible] factors might have recommended [illegible] research not be undertaken. Ou[illegible]stence resulted from the belief tha[illegible] regarding contextual stimuli cannot [illegible] these complexities, so we chose to acc[illegible]hem in the initial experiment. Furthermore, in the analysis of the data, attem[illegible] have been made to determine if the a[illegible]mentioned interference effects were in [illegible] operating during the single study tria[illegible]

Experiment I

Method

The *S*s were given a single study [illegible] 36-pair list prior to being tested. Before [illegible]t was presented, they were told that their [illegible] would be tested in 2 ways. First, resp[illegible] term recall would be required; and second, a fo[illegible]d-choice recognition test would be given, in whic[illegible]ets of 2 pairs would be shown with the request to choose the correct pair (the pair presented on the study trial) in each set. Four study-test trials were used. Half of the *S*s were tested in each of the following orders: recall, recognition, recognition, recall; and recognition, recall, recall, recognition. The *S* did not know which test he would be given for a trial until after the study trial. The major interest was in the memory for the items after the first trial. Analysis of scores beyond the first trial has provided no useful information beyond that given by the first trial, so that all evidence to be presented is derived from the first-trial performance.

Lists. The 7 lists were constructed from 36 3-letter words of minimal letter duplication and from 36 2-syllable words having concreteness ratings of 6.0 or higher in the Paivio, Yuille, and Madigan tables (1968). The 3-letter words served as stimulus terms, the 2-syllable words as response terms. The instructions prior to learning included illustrations of both types of words (stimulus terms and response terms) so that *S* knew what to expect on the 2 types of retention tests.

The 7 lists represented the context manipulation, and these conditions will be designated by the number of stimulus terms used in the lists. For Condition 1, a single stimulus term was used, so it was essentially like a true free-recall list in that the stimulus terms could not be used as a differential

cue. In successive lists there were 2, 3, 4, 9, 18, and 36 different stimulus terms. For Condition 36, the 36 3-letter words and the 36 2-syllable words were paired randomly. When the number of stimulus terms was less than 36, they were chosen randomly from the 36 and paired randomly with the response terms, subject to the restriction that each was paired an equal number of times. Response terms paired with a common stimulus term were blocked within the list. Thus, in Condition 9, for example, there were 9 blocks of 4 pairs each, with the same stimulus term used within each block. The ordering of the blocks was random.

Procedure and testing. The data were collected by a group procedure. Prior to testing, each group was assigned randomly to 1 of the 7 lists. Within each group, the recognition test and the recall test were assigned randomly for the first trial after the study phase. For the recall test *Ss* were simply asked to write down all of the response words (which had been identified in the instructions as the right-hand 2-syllable words). For the recognition test, *Ss* were given a test sheet on which there were 36 sets of 2 pairs each, both pairs in each set having the same response term but a different stimulus term. The *Ss* marked the correct pair in a blank provided for this purpose. Both stimulus terms were always from the list. In Condition 2 there were only 2 stimulus terms, so these occurred in all pairs. For all other conditions, the incorrect stimulus term was chosen randomly, subject to the restriction that it was not used a second time until all other stimulus terms had been used once. The ordering of the sets of pairs on the recognition test sheet was random with respect to the order of the pairs on the study trial.

The pairs were presented for study at a 4.5-sec. rate. Approximately 3.5 min. were allowed for the test period. Small groups of *Ss* were tested until each condition had a minimum of 20 *Ss* each. Seven of these groups of 20 *Ss* had been tested for recall on the initial trial, 6 for recognition. The list designated Condition 1, in which only one stimulus terms was used, could not be administered as a recognition test in a meaningful manner. The *Ss* given this list had 4 successive recall trials.

Results

Response-term recall. The percentages of items recalled for each of the seven conditions are plotted in Figure 1. The base line is described as the log of the number of different stimuli involved in the 7 lists. Thus, the number of different contextual changes increases as the number of different stimulus terms increases. In general, recall decreased as the number of contextual stimuli increased, but the relationship is by no means a regular one. The differences among the recalls of the 7 groups are

FIG. 1. Response-term recall as a function of number of different stimuli for the 36 response terms.

statistically significant, $F(6, 133) = 4.33$, $p < .01$. Performance under Conditions 3 and 4 is less than that for Condition 9. However, the statistical reliability of these differences, as well as of the difference between Conditions 1 and 2, cannot be demonstrated. For the moment then, it may be said that as the number of different stimulus terms increased, there was a decrease in the number of response terms recalled. Stated in another way, the relationship shows that as the number of different words presented (stimulus words plus response words) increased, recall decreased. This might be interpreted in terms of memory load, although the data to follow make it unlikely that this interpretation will suffice.

Position and block effects. Figure 2 provides an initial inspection of recall as related to the position of the item on the study trial. The 3 lines represent different portions of the list, viz., primacy (Items 1–6), recency (Items 31–36), and all items in the body of the list (Items 7–30). First, it can be seen that there was a general decrease in recall of items from the body of the list as the number of different stimulus terms increased. Next, there was a lack of primacy and recency effects for Conditions 3 and 4; that is, recall for the initial and final positions did not differ appreciably from recall of the items within the body of the list. All other conditions show primacy and recency effects, although

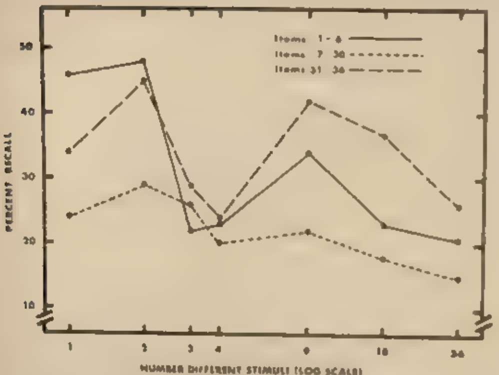


FIG. 2. Response-term recall as related to primacy (Items 1-6), recency (Item 31-36), and all other items (7-30), and to number of different stimuli.

the primacy effects are relatively slight for Conditions 18 and 36.

It was pointed out earlier that when *S* tries to learn response terms to a common stimulus, interference might occur. This in turn could lead to unlearning. If unlearning occurs, items early in a block having the same stimulus term should be more poorly recalled than items occurring late in the block. The data have been examined to see if they give any support to this notion. For Condition 18, where there were only 2 response terms for each stimulus term, recall of the first and second items (combined across the 18 blocks) was essentially equal. For Condition 9, very clear position effects were apparent within the blocks. The first item within the block showed the best recall, and the second was recalled better than the third and the fourth, the latter 2 being about equal. This finding provides no support for unlearning within blocks. Rather, it appears that each new stimulus term served to define a new rehearsal block, and as *S* proceeded through the 4-pair block, rehearsal fell most heavily on the first item in the block. In Condition 4 (9 response terms for each of 4 stimulus terms) there were no within-blocks differences of any consequence. The position effects within blocks were indistinguishable from those of Condition 1 when the latter was scored comparably.

For Conditions 2 and 3, the picture changes again. The percentages correct for 6 successive blocks of 6 items which have been determined for the 2 conditions. In Condition 3, each stimulus was paired with 12 response terms, hence each block with a single stimulus term consisted of 2 successive sixths. For Condition 2 the first 18 items had a common stimulus term and the second 18 had a different common stimulus term. Plotting in sixths means that the first 3 sixths had the same stimulus term, and the last 3 sixths the second stimulus term. Figure 3 shows that in Condition 3, performance increased for each successive block of 12 items and within each block, with the last 6 items recalled more poorly than the first 6. Thus, the within-blocks effects are much like those for Condition 9, and again, it would be concluded that no unlearning is apparent within blocks. For the first block of 18 items in Condition 2 there is a sharp decline across the 3 sixths. This is followed by a sharp increase as the stimulus term changes, followed again by a decrease, with what is apparently a recency effect in the final block. For both conditions represented in Figure 3, it becomes particularly clear that a change of stimulus context had a marked effect on the specific items recalled. Finally, for Condition 1, the recall by position was much like that for Condition 2 for the first 18 items, but, of course no marked increase in

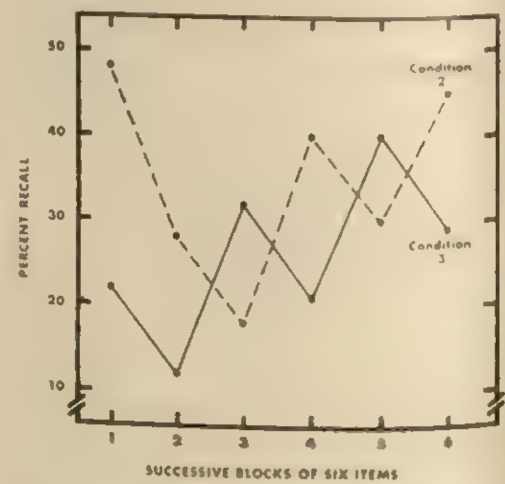


FIG. 3. Response-term recall for successive blocks of 6 items with 2 and 3 different stimulus terms.

recall occurred for the fourth sixth as was true for Condition 2.

Recognition. In the recognition tests, *S*s were required to choose between 2 pairs in 36 sets, one pair in each set being correct, the other incorrect. The percentages correct for the 6 conditions are shown in Figure 4. Unlike recall, recognition performance generally increased as the number of stimulus terms increased. The difference overall is statistically reliable, F (5, 114) $= 2.49$, $p < .05$, although the maximum difference is only about 10%.

Primacy effects, although small, were present in the recognition scores, but there were no recency effects. Furthermore, differences within blocks, so evident in several of the conditions for recall, were much less apparent in the recognition data. Correct decisions made on pairs in recognition might be expected to be correlated with the recall of the response terms of those pairs when summed across *S*s. For the 6 conditions, the product-moment correlations across the 36 items varied between .01 and .31. Although even the largest value is not significant by conventional standards, the fact that all 6 correlations were positive indicates a slight relationship, perhaps reflecting the common primacy effect found in both recall and recognition. In any event, these data indicate that knowledge of associations, as determined by recognition of appropriate pairing, holds little relationship to the recall of the response terms of those pairs.

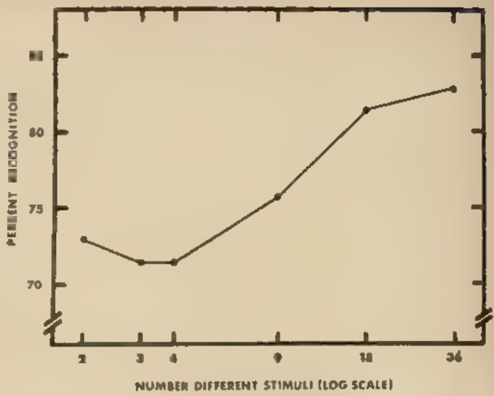


FIG. 4. Pair recognition as a function of number of different stimuli during the learning trial.

DISCUSSION

Discussion

As pointed out in the introduction, it had been expected that learning of the different lists would produce a rather complicated picture, and this expectation seems to have been realized. Yet evidence for within-lists interference produced by the A–B, A–C paradigm is not easy to identify. The recognition data do suggest that the fewer different response terms associated with a given stimulus term, the more intact the association. There was no support for the idea that unlearning might occur within blocks. Except for Condition 4, within-blocks differences were quite marked, with performance on the initial items in a block being better than on later items. One interpretation that has already been suggested is that the blocks, at least the smaller ones, provide a ready rehearsal set. This interpretation is not without some ambiguity, since the same type of effect would be expected if the early pairs interfered with the learning of the later pairs. However, as will be seen in Experiment II, it is likely that the interference component is relatively small.

For some lists (Figure 3) there was a sizeable increase in recall for each successive block. This could be interpreted as retroactive interference of later blocks on earlier blocks. Such interference would have to result from an A–B, C–D paradigm. This may not be plausible. An alternative interpretation is that the increase in recall across blocks simply represents differences in recall associated with retention intervals of different lengths.

Context can play a role in recall only if it is reinstated at the time of recall. There is little doubt that for the lists with relatively few stimulus terms, these terms could have been recalled. The number of response terms recalled was directly related to the likelihood that the stimulus terms could be recalled. Yet, it should be clear that there is no convincing evidence that recall of the response term was mediated by the recall of the stimulus term. So far as we can tell, there is no way to design an experiment to make such a determination directly. Although not reported above, clustering by stimulus terms occurred only in Conditions 9 and 18 in amounts greater than could be expected for clustering mediated by position alone (as determined from Condition 1). This makes it likely that stimulus-

term recall preceded response recall, but such a result does not demand this interpretation.

It is apparent that the complexity of the findings requires more analytical work if the processes associated with context shifts are to be untangled. Experiments II and III were viewed as first steps toward this untangling, with the focus being on the interference (and unlearning) for the within-lists A–B, A–C paradigm.

Experiment II

In Experiment II the A–B, A–C pairs have been unblocked. This was viewed as a means of breaking up, or at least minimizing, the possibility that common stimulus terms may be used to define rehearsal blocks. There were 3 lists, corresponding to those in Conditions 3, 4, and 9 in Experiment I. As a second variable, half of the *S*s were given the stimulus terms at the time of recall of the response terms, and half were given free recall of the response terms. All *S*s were instructed that they would be tested for response-term recall with and without the stimulus terms present. Thus, the instructions again met the 2 requirements believed necessary if the stimulus terms were to simulate contexts.

Method

The pairs used in Experiment I for Conditions 3, 4, and 9 were assigned randomly to positions in the list, subject only to the restriction that pairs with the same stimulus were not allowed to occupy adjacent positions. Three groups of 20 *S*s each were given response-term recall (as in Experiment I), and 3 additional groups of 20 *S*s each were given the stimulus terms on the recall sheet and were asked to pair the response terms with the appropriate stimulus terms. The *S*s in all groups were instructed that on successive trials they would be given both free recall and cued recall, the particular type of recall to be identified after the list had been presented for learning. As with Experiment I, trials beyond the first are not included in the present analysis, these extra trials being given merely to fulfill the obligation implied by the instructions. All other procedures were the same as those in Experiment I.

Results

Recall. The recall for the 6 groups varied between 17.2% and 23.9%. Mispaired responses for the cued condition were counted as correct. Neither condition, as defined by number of stimuli (F = [illegible]), nor cuing vs. noncuing ($F = 1.61$), was significant source of variance. An analysis of the free recall, which included the data from the 3 conditions in Experiment I, showed no statistical difference, although the lack of a difference is difficult to interpret as the 2 experiments were performed at different times in the school year.

Position and recall. There were no systematic differences in recall as a function of position of the items on the same trial between cued and free recall. Therefore, the data have been combined to evaluate the influence of input position on recall. Some general observations will be made initially. All 3 conditions showed marked primacy and recency effects, in contrast to Experiment I, where Conditions 3 and 4 showed little evidence of such effects. For the 3 conditions of Experiment II, primacy (Items 1–6) increased directly as number of stimulus terms increased, the values for recall being 28%, 35% and 45% for Conditions 3, 4, and 9, respectively. Recency effects (Items 30–36) were greater for Condition 3 than for Conditions 4 and 9, the values being 33%, 25%, and 26%, respectively. For items in the body of the list (7–30), recall decreased by 8% between Condition 3 (20%) and Condition 9 (12%), with Condition 4 intermediate.

The major interest was in the recall of successive response terms having the same stimulus term. There is no satisfactory way to remove the overall primacy and recency components from this analysis, since buffer items were not used. In fact, these effects should probably not be removed, since they may reflect in part the interference which is the focus of the analysis. The approach used was to examine recall as a function of position in the series of items having the same stimulus term and to compare this recall with the results of Experiment I. In Experiment I the pairs were blocked by stimulus terms, whereas in Experiment II the pairs with the same stimulus term were randomized throughout the 36 positions. The data were summed across different stimuli for the first occurrence of the stimulus terms,

the second occurrence, and so on. For Condition 9, each stimulus term occurred 4 times, hence each of the 4 positions was represented by the 9 different stimulus terms. For Condition 4 each stimulus term occurred 9 times, therefore, each of the 9 positions was represented by 4 different stimuli. Finally, for Condition 3, each stimulus occurred 12 times, so each of the 12 positions was represented by 3 different stimulus terms. The percentage recall was determined for each position in the sequence for each of the conditions of the 2 experiments. These values are plotted in Figure 5.

The first finding to be noted is the initial drop which occurs, this being greater for Experiment II than for Experiment I. The second finding of note is that following the initial drop, recall in Experiment II tended to increase across successive positions, whereas no such trend was apparent in Experiment I. This was particularly clear for Condition 3, where each stimulus occurred 12 times. The data for Condition 3 may be related to the plot for Experiment I, shown in Figure 3. For each of the thirds in Figure 3, the first 6 items within a block were better recalled than the second 6 items. This is reflected in Figure 5 for this condition, where recall for the first 6 positions (31%) is greater than recall for the last 6 positions (21%). In contrast, in Experiment II recall for the first 6 positions (20%) is less than for the last 6 positions (27%).

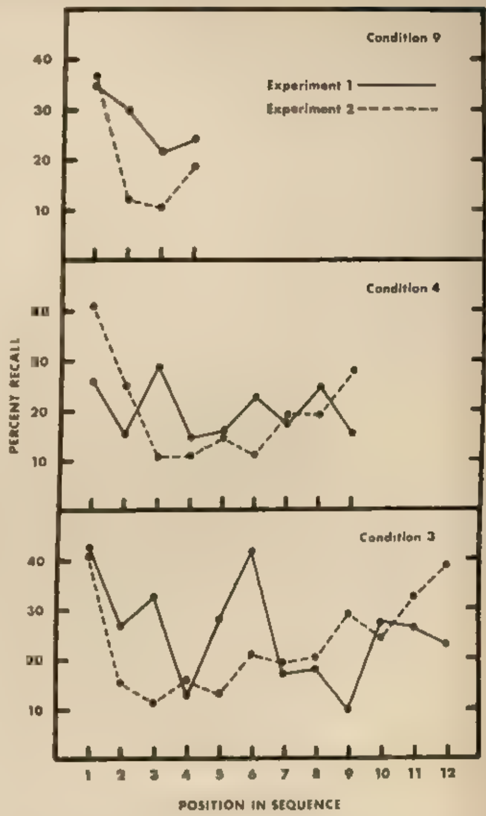


FIG. 5. Recall of response terms having the same stimulus terms as a function of position in the sequence. (In Experiment I the pairs with the same stimulus terms were blocked, while in Experiment II they were not. The conditions with 3, 4, and 9 different stimulus terms are represented in the 3 panels.)

Discussion

The results in Figure 5 for Experiment II suggest that unlearning occurred within the list when the same stimulus term was used for successive response terms. The first pair seems to be "protected" from unlearning, but beyond this there is a clear trend for each successive response term to be recalled better than the preceding ones, although the second position is somewhat ambiguous. Increasingly better recall with each successive stimulus should happen if each successive pairing produced more and more unlearning of previous items.

There are undoubtedly many other explanations in addition to unlearning which might be offered to account for the gradual increase in recall across positions (e.g., shorter retention intervals for later items). If unlearning is occurring, one expectation might be that absolute recall should have been better on, say, Positions 3 and 4 for Condition 9 than for Condition 3. This might be expected on the grounds that in the latter condition there were more successive items to produce unlearning than in the former. However, a complicating factor may arise from different amounts of retroactive interference between items with different stimuli. This possibility is greater in Condition 9 than in Condition 3.

When pairs are blocked, as in Experiment I, it seems likely that the blocking provides a ready rehearsal set which would counter unlearning, but which would lead to poorer recall for items late in the block than for those early

in the block. The purpose behind the unblocking in Experiment II was to remove the ready-made rehearsal blocks. It was believed that evidence for unlearning might be obtained under these circumstances. If unlearning did occur, as presumed in Experiment II, spontaneous recovery might be expected. Since unlearning is assumed to be greater for items in the initial positions, spontaneous recovery should be greater for the items in these positions than for items in the late positions sharing the same stimulus term. The purpose of Experiment III was to determine if spontaneous recovery could be detected in a delayed retention test.

Experiment III

There is no evidence as to the course of recovery over time following free-recall learning. Therefore, a 20-min. interval was chosen on the basis that spontaneous recovery has been observed within this range when the A–B, A–C paradigm holds between lists (e.g., Postman, Stark, & Henschel, 1969).

Method

A single list of pairs was used, this being the list identifying Condition 3 in the first 2 experiments. There were only 2 basic conditions, an immediate cued recall (the same as given in Experiment II), and for another group of *S*s, a cued recall after 20 min. Two random orders of the 36 pairs were employed, but since the results showed that [illegible] were equivalent, no further mention will be made of this factor. Each group was made up of 34 *S*s. The *S*s were instructed prior to the presentation of the list that they would be tested for recall and that the response terms were to be listed under the appropriate stimulus term. The *S*s were tested individually, and the words were presented at a [illegible] rate. Those *S*s given the 20-min. recall worked on a pyramid puzzle during the interval

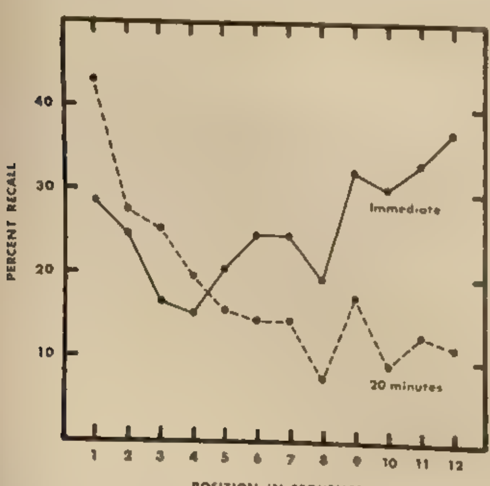


FIG. 6. Response-term recall as a function of the length of the retention interval (immediate or 20 min.) and position in sequences of response terms having the same stimulus term.

Results

The immediate recall was 25.4%, slightly higher than that for the comparable condition in Experiment II (23.9%). The group given recall after 20 min. produced 18.4% of the response terms.

The critical data are shown in Figure 6. The plot corresponds to that for Figure 5. Looking first at the immediate recall, it can be seen that except for the 2 initial positions, the recall by position was much the same as in Experiment II and as displayed in the lower panel of Figure 5. Why the primacy effect was greater in Experiment II than in Experiment III is not known. Recall by position after 20 min. was quite different from the corresponding curve for immediate recall. For the 4 initial positions, recall was actually higher after 20 min. than on the immediate test. On all subsequent positions it was less. In short, the results are not unlike those to be expected if unlearning and spontaneous recovery occurred. The interaction between position and length of retention interval is highly reliable, F (11, 726) = 11.36, $p < .01$, using an arc sine transformation of the data. Evaluated independently, recall for the first 4 positions is statistically greater for the 20 min. group than for the immediate group, F (1, 66) = 4.01, $p < .05$.

If there is output interference and if *S*s in the 20-min. group gave priority to the recall of the early items, while those with immediate recall gave priority to the recall of items late in the list, the results shown in Figure 6 might be produced. However, an analysis showed no difference between output order for the 2 groups. The data appear to be consonant with the notion that spontaneous recovery occurred and that

this retarded the forgetting of items. The greatest retardation occurred early in the list, where the greatest amount of unlearning is presumed to have taken place. Taken at face value, the data indicate that recovery even occurred for the first pair in the sequences.

General Discussion

The procedures of the present experiments were intended to simulate context shifts or changes. In the first experiment, recall was inversely related to the number of context shifts, although this relationship was neither strong nor regular. Number of context shifts was directly related to number of different stimulus terms. The relationship between context shifts and recall may be complicated by internal interference when the same context serves as a stimulus term for more than one response term. The fact that recognition showed that the association between stimulus and response terms remained more stable with increasing numbers of stimulus terms suggests that interference may counteract to some extent the positive effect of a relatively few context stimuli. Nevertheless, the positive and negative factors which seem to be operating jointly to produce recall performance may be quite realistic and applicable when notions concerning the role which context plays in free-recall learning are advanced. This is to say, simplified theoretical notions of the role of multiple contexts may not be appropriate.

As a context shift occurred—as pairs with the same stimulus term were followed by a pair with a new stimulus term—performance on the initial pairs after the shift was markedly enhanced under some conditions. When the pairs of words with the same stimulus terms were randomized within the list rather than being blocked, the learning pattern within the sequence of response words with the same stimulus terms changed. It is possible conceptually to think of contexts, particularly internal contexts, which may be present for only one word, and then reoccur at a later point in the list for a new word. Experiment II simulated this reoccurring context. The evidence from this experiment was interpreted as showing unlearning for pairs having the same context, with the earlier occurring pairs being unlearned as a consequence of the occurrence of the later pairs. Experiment III gave some support to this interpretation by showing that spontaneous recovery of unlearned associations may occur over a retention interval. Such data again suggest that contextual stimuli may produce a rather complicated set of relationships.

References

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76** (1, Pt. 2).

Postman, L., Stark, K., & Henschel, D. Conditions of recovery after unlearning. *Journal of Experimental Psychology Monograph*, 1969, **81** (1, Pt. 2).

Turvey, M. T., & Egan, J. Contextual changes and release from proactive interference in short-term verbal memory. *Journal of Experimental Psychology*, 1969, **81**, 396–397.

(Received April 28, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 142–150

DICHOTIC LISTENING AND SEQUENTIAL ASSOCIATIONS IN AUDITORY SHORT-TERM MEMORY [1]

CATHERINE G. PENNEY [2]

Memorial University of Newfoundland, St. John's, Newfoundland, Canada

Two experiments are reported in which a dichotic (bisensory) stimulus presentation was used. In one experiment, 4 stimulus pairs were presented in a trial, 2 dichotic pairs being alternated with 2 visual pairs. The preferred recall order was mode by mode when all words were to be recalled, and ear by ear for auditory words. In Experiment II, 8 words were presented in a trial, but 4 words, 2 auditory and 2 visual, were simultaneous. The *S* was instructed to recall either the word presented before or after the probe but in the same ear or same part of the screen, or the word presented at the same time and in the same modality as the probe. The results were interpreted as evidence for the existence of strong associations between successive auditory stimuli.

In dichotic listening, auditory stimuli are presented 2 at a time, one to each ear, and *S* is asked to recall. If *S* is not instructed to report the items in any particular order, he tends, at fast presentation rates, to report all those items presented to one ear and then those presented to the other ear (Broadbent, 1954, 1956). At slow rates, he generally reports the first pair of items, then the second pair, and then the third pair.

Broadbent (1958) interpreted this "ear-by-ear" order of report in terms of the difficulty of switching attention between channels of information. He hypothesized that a selective filter mechanism can be set to select 1 channel, distinguished on the basis of some physical characteristic, and to reject all other channels. Because of the time involved in switching channels, ear-by-ear order of report is more efficient at fast, but not at slow, presentation rates. Subsequent work has indicated that at least 1 main assumption of Broadbent's model, i.e., that only simple physical characteristics could be analyzed prior to the filter, is incorrect. Several studies (e.g., Gray & Wedderburn, 1960; Yntema & Trask, 1963) have indicated that meaningfulness of input items could define a channel, and that meaning therefore must be decoded prior to or at the level of the filter.

Yntema and Trask (1963) suggested that *S* can perceive both items presented simultaneously and can tag items according to either their meaning or ear of arrival at input. The presence of the tags provides *S* with a way to systematically retrieve the items. With homogeneous stimuli, ear of arrival provides the basis for output organization, but with distinct classes of stimuli, recall by stimulus type is apparently preferred. It is not clear, however, why time of arrival is not equally useful.

Contrary to Yntema and Trask (1963), Bryden's (1964) data show that in some cases meaning alone is not a sufficient basis for a retrieval plan. He presented highly associated pairs of words where 1 word of each pair was presented to each ear. At fast presentation rates of 2 pairs/sec, he found that *S*s preferred to recall in an ear-by-ear order and not, as might be expected from the findings of Yntema and

[1] This study is based on a doctoral dissertation submitted to the University of Toronto in 1971. The research was supported by Grant APA-146 from the National Research Council of Canada and Grant OMHF-164 from the Ontario Mental Health Foundation to Bennet B. Murdock. The author would like to thank Bennet Murdock, Robert Lockhart, and Joan Foley, who served on the thesis committee; and Derek Corcoran and Ted Rowe, who read earlier drafts of the manuscript and offered many helpful suggestions. A short version of this report was read at the meeting of the Canadian Psychological Association, Montreal, Canada, June 1972.

[2] Requests for reprints should be sent to Catherine G. Penney, Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

Trask, in the terms of the associated pairs. Whitmore (1972) presented words dichotically in such a way that if *S* reported items in temporal pairs, he would have a meaningful sentence. Without specific instructions to the contrary, *S*s tended to report in an ear-by-ear order. It seems that Yntema and Trask's retrieval explanation is not entirely adequate, as *S*s appear to freely report according to meaning only when the meaningfully related items are not simultaneous.

A further complication is presented by the work of Savin (1967) and Rollins, Everson, and Shurman (1972). When 2 digits are presented simultaneously over a loudspeaker and do not differ in their physical characteristics, *S* presumably has no basis on which to channelize the information. One would therefore expect the items to be recalled in pairs. However, Rollins et al. and Savin found that *S* still preferred to recall the items in a sequential order, that is, an item from the first pair, one from the second pair, and the remaining item from the first pair and the last item.

When, instead of the 2 ears, the sensory modalities of audition and vision define the channels used, the picture becomes even more complex. The *S*s prefer to recall 3 audiovisual pairs by modalities at fast presentation rates, and as pairs at slow rates (Madsen, Rollins, & Senf, 1970). Output by modalities also results in higher recall than output by pairs (Broadbent & Gregory, 1965).

However, other findings indicate that the dichotic and bisensory situations are not strictly analogous. In the bisensory analogue of the Yntema and Trask (1963) study, Madsen et al. (1970) found that recall by type of stimuli resulted in a low level of performance at both fast and slow presentation rates. Broadbent and Gregory (1965) also found that recall according to types was worse than recall by modalities. Moreover, when auditory stimuli are presented alternately to the 2 ears, *S*s have no difficulty in recalling them in their sequential order (Moray, 1960). This finding does not hold when stimuli are presented alternately to the ear and eye (Broadbent & Gregory, 1961).

Finally, in contrast with auditory studies, an examination of the visual analogue of dichotic listening shows that the channel-by-channel report is not a preferred output order for recall of simultaneous stimuli. When separate digits are presented to each eye (Sampson & Spong, 1961) or 2 superimposed digits are presented to both eyes (Schurman, Everson, & Rollins, 1972), the preferred recall order is in terms of temporal pairs. However, Schurman et al. showed that *S*s were equally accurate with a temporal-pairs or a sequential recall order.

In summary then, a channel-by-channel recall order is obtained with both a dichotic and bisensory stimulus presentation, but not with a visual presentation. The variables of type of stimulus material, associative relations between items, and whether items on the 2 channels are simultaneous or successive do not have similar effects on recall order for the dichotic and bisensory task. These findings suggest the need for further investigation of the differences between audition and vision as input channels. These differences were explored in the present experiments using a task that combines dichotic and bisensory presentation. In Experiment I, pairs of dichotic auditory stimuli were alternated with pairs of simultaneous visual stimuli, and *S* recalled either the auditory stimuli, the visual stimuli, or both. With this type of stimulus presentation, one would expect recall to be organized by modalities, with either a temporal-pairs organization within each modality, or no strong organization at all.

Experiment I

Method

Design. On each trial, 8 stimulus words were presented as 4 successive pairs. On half of the trials the first and third pairs were presented dichotically, while the second and fourth pairs were visual (AVAV). On the other trials, the order of the word pairs was visual, auditory, visual, auditory (VAVA). At the end of each list, a cue was presented, indicating 1 of 3 recall conditions: recall of all of the words, only the auditory, or only the visual.

For each of these recall conditions, the postlist cue was auditory on half of the trials and visual on the other half. The 12 treatment combinations (obtained by factorially varying the cue modality, the recall instruction, and the order of the auditory and visual word pairs) were presented in a random order within a session with the restriction that each condition was tested once in each block of 12 trials. There were 6 blocks of experimental trials, giving 6 replications of each treatment combination for each *S*. To ensure that the results obtained would not be limited to a specific set of stimuli, 2 different sets of stimulus material were constructed. Six *S*s were tested on each set.

Subjects. The *S*s were first-year male and female University of Toronto students who participated in psychology experiments for credit. All *S*s were native speakers of English, were aware of no hearing impairment, and had normal or corrected vision. Three male and 3 female *S*s were individually tested on each set of stimulus materials.

Stimulus material. For each set of stimulus materials, words were randomly selected without replacement from a pool of 1,430 1-syllable common English words of all grammatical classes. The visual words were typed on a roll of plain white computer paper and presented over closed-circuit television. To construct dichotic tapes, a male and a female graduate student were isolated in soundproof booths which contained a microphone and a device set by *E* to click at a .75-sec. rate. The speakers uttered the stimulus words in time with every second click, starting after 4 or 5 warning clicks. The clicks were recorded on the tape to provide *S* with a signal before the start of each trial. The auditory and visual stimuli were synchronized by means of a Uher diapilot and a Hunter timer. The timer caused a video control monitor to transmit the signal to *S*'s television monitor for .75 sec. Except when a stimulus was being shown, *S*'s screen was bright but blank. The postlist cue was timed to occur 1.5 sec. after the start of the last stimulus pair.

Procedure. The *S* was given a sheet of instructions that described the simultaneous presentation of 2 words in 1 sensory modality and the alternation of the 2 modalities. The single word at the end of each list served as a recall cue. This last word was *one*, *two*, or *three* and was presented either auditorily or visually. The *S* was instructed to ignore the modality of the cue but to attend to the number. *One* signaled recall of all words, *two* indicated only auditory words, and *three*, the visual words. The *S* could recall the words in any order and was instructed to guess when uncertain.

The *S*'s performance was monitored during 12 practice trials to ensure that the instructions were followed. If *S* experienced any difficulty, a few extra practice trials were given. No *S* was rejected for inability to do the task. After the practice trials, any questions pertaining to the procedure were answered, and the 72 experimental trials then followed.

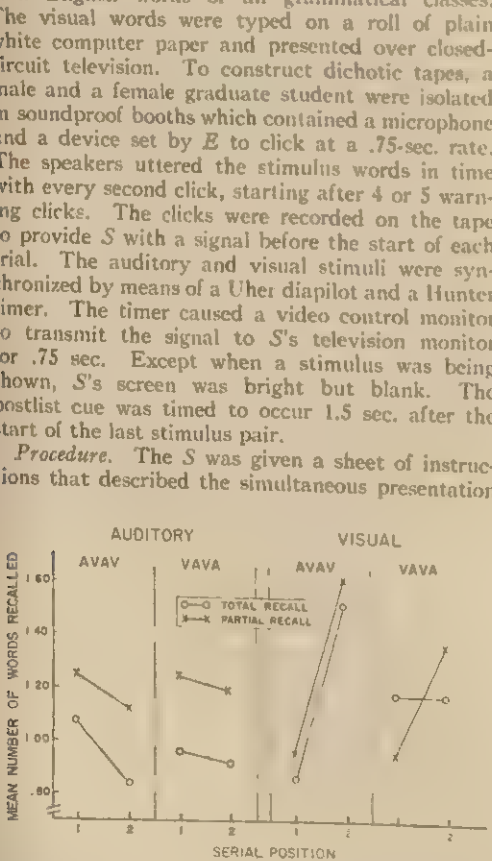


FIGURE 1. Recall of auditory and visual words as a function of serial position, modality order (A = auditory, V = visual), and type of recall trial.

Results

The criterion for scoring was lenient, and a different criterion was used for auditory and visual items. The reason for this was that, particularly for auditory items, there was a large number of what were possibly auditory perceptual errors. For example, *S* might report *block* or *blot* when the recorded word was *plot*. In an earlier experiment (Penney, 1973, Experiment I) with a similar stimulus presentation, a comparison of strict and lenient scoring criteria showed the pattern of results to be the same for both, with the main effect of using the lenient criterion being the raising of the level of recall of auditory items to approximately that of the visual items. The scoring criteria used were as follows. An auditory item was counted correct if the word spoken by *S* had no more than 2 incorrect phonemes. A visual word was counted correct if a possible spelling of *S*'s spoken recall contained no more than 1 incorrect, omitted, or transposed letter. For example *ace* would be accepted for *age* or *through* for *though*.

Figure 1 shows the mean number of words recalled from the pair as a function of serial position and modality. The separate panels are for the AVAV and VAVA presentation orders. The 2 curves in each panel represent the total- and partial-recall trials.

A $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 6$ analysis of variance was conducted with the follow-

ing factors: modality (auditory or visual), serial position within a modality (first or second pair), type of recall trial (total or partial), modality of the recall cue, order of the auditory and visual word pairs (AVAV or VAVA), set of stimulus materials, and *S*s. The dependent variable was the proportion of words recalled out of the 6 replications. As arc sine transformation was applied to these proportions (Murdock & Ogilvie, 1968).

Because of the different scoring criteria for auditory and visual items, a direct comparison of auditory and visual recall cannot be made. However, the interactions with presentation modality are independent of the main effect and can be interpreted. Figure 1 indicates that there is very little difference in recall as a function of serial position for auditory items. For visual stimuli, on the other hand, words from the second pair are generally recalled much better than those from the first pair. The serial-position effects are smaller for the VAVA order, with a flat curve in the total-recall condition. The Serial Position × Modality interaction was significant, $F\ (1, 5) = 20$, $p < .01$. These serial-position effects were large enough to cause the main effect of serial position to be statistically significant overall, $F\ (1, 5) = 8.9$, $p < .05$. There were, however, no reliable differences in recall of auditory and visual stimuli.

Recall of the visual words in the last serial position was lower in the VAVA condition than in the AVAV condition, but recall of the first pair was slightly better. Recall of auditory words did not seem to be affected by order. The Serial Position × Order interaction was significant, $F\ (1, 5) = 63$, $p < .005$, and the Serial Position × Modality × Order interaction was also significant, $F\ (1, 5) = 17$, $p < .01$.

Considerably more auditory words were recalled on partial-recall trials than on total-recall trials. For visual words, the effect of the type of recall trial was less marked. Overall, the main effect of the type of recall trial was highly significant, $F\ (1, 5) = 75$, $p < .005$, and the Modality × Type of Recall Trial interaction was marginally significant, $F\ (1, 5) = 6.6$, $p < .05$.

The total-recall trials were further examined to discover the order of recall of the different modalities. On 198 of 288 total-recall trials (68.8%), *S* recalled some words from both modalities and switched between modalities during output only once. On 44 of the remaining 90 trials, *S* recalled words from only 1 modality. Thus, there is still a strong preference for modality-by-modality recall even when auditory and visual stimuli are presented alternately (Broadbent & Gregory, 1965).

If only those total-recall trials on which *S* made exactly one modality switch are considered, 77.3% of the time *S* recalled the visual words first. The output priority given to visual items on total-recall trials explains why there is so little difference for visual stimuli on total- and partial-recall trials. The main effect of cue modality was not significant, nor was any interaction with cue modality.

The data of primary interest in this experiment are those concerning the order of recall of the auditory stimuli. Only partial-recall trials were examined. The first 2 words recalled were classified as having been presented at the same time, in the same place, or neither. For visual words, the "same-time" recall order occurred on 57.3% of the trials. A "same-place" output order was found on only 10.8% of trials. For auditory words, on the other hand, the first 2 words recalled had been presented to the same ear on 39.9% of the trials and at the same time on 21.2%. A sequential output order, in which the first 2 words recalled were presented to different ears and at different times, occurred on 18.1% of the trials. There is still a substantial preference for a sequential recall order over a temporal-pairs output order even when there are visual stimuli intervening between subsequent items presented to the same ear.

The finding of a preference for a sequential order of report in the present experiment would seem to pose a problem for Broadbent's (1958) model. A strategy of

attending to only one ear is unlikely, as on 131 of 288 auditory partial-recall trials, *S* was able to recall 3 or 4 auditory words. Of these trials, an ear-by-ear order of report occurred on 50.4%, a temporal-pairs order on 32.06%, and a sequential order on 17.56%.

In fact, many *Ss* reported that they tried to attend to the visual stimuli during presentation, as they found these more difficult to remember, and they certainly recalled them first on total-recall trials. If the auditory stimuli actually constituted the rejected channel, as *Ss*' introspective reports would suggest, the finding of an ear-by-ear report order is puzzling. The auditory items must stay in the sensory buffer while the visual information is passed through the limited-capacity system. When *S* is free to switch attention to the auditory mode, the most logical strategy would seem to be to process the 2 oldest items first.

It is clear that Broadbent's (1958) model, even in its revised form, has difficulty with these data. Equally, Yntema and Trask (1963) provided no explanation for the fact that *Ss* do not use time of arrival as a basis for organization. In the present experiment, time would seem to be as good as or better than ear of arrival as a basis for a retrieval plan. However, it may be that under different retrieval conditions, time of arrival could be an effective cue. In the second experiment, a partial- rather than a total-report procedure was used in order to see if *S*'s encoding and retrieval strategies were affected.

Experiment II

Most experiments on dichotic split-span memory have used a whole-report procedure and measured either the relative efficiency of, or *Ss* preference for, the ear-by-ear or temporal-pairs output order. In the second experiment, a partial-report technique was used, with *S* reporting only 1 word on each trial. The probe was either the word that had occurred at the same time as the to-be-remembered (TBR) word but in the other ear and voice (same-time probe) or a word that had been presented either before or after the TBR word in the same ear and voice (same-place probe). The requirement in earlier studies that *S* report every item on each trial may have necessitated, as Yntema and Trask (1963) suggested, some kind of tagging according to ear of arrival or stimulus type in order for *S* to have an organized retrieval plan. This tagging strategy may not be necessary when a probe technique is used. When *S* knows in advance of stimulus presentation whether he is to receive a same-time or same-place probe, he should be able to encode the items accordingly. Therefore, one would expect no difference in overall recall in the 2 probe conditions. The occurrence of any difference in recall would suggest that *Ss* are not always able to optimally encode information in this situation.

Method

Design. The second experiment involved a slight modification of the stimulus-presentation procedure used in Experiment I, in that the pairs of auditory and visual stimuli were presented simultaneously. The *S* was still presented with 8 words on each trial, but 4 rather than 2 words were presented at a time. The probe word was randomly selected from the 8 words with the restriction that in each block of 8 trials each of the 8 positions was tested once. There were 8 blocks of experimental trials and 2 blocks of practice trials, giving a total of 80 trials. All *Ss* were tested on both the same time (ST) and same place (SP) conditions, half receiving the ST condition first and half the SP. Two sets of stimulus material were used. Of those *Ss* receiving the ST condition first, half had Set A for this condition and Set B for the SP condition, while the remaining *Ss* had Set B for the ST condition and Set A for the SP condition. In this way, there was no confounding of probe conditions, stimulus set, and practice effects. The design is therefore a fully crossed factorial with position tested, replications (blocks), and instruction conditions varied within *Ss*, and presentation order of the instruction conditions and of the stimulus materials varied between *Ss*.

Subjects. Eight male and 8 female University of Toronto graduate students with normal or corrected vision and no known hearing defects were tested individually and were paid $4 for the 2 sessions.

Stimulus materials. Two sets of stimulus materials were used. For each set, 640 words were randomly selected from the word pool used in Experiment I. Recording of auditory stimuli and presentation of visual stimuli were similar to Ex-

periment I. The presentation rate was 1 set of 4 words per sec., with the actual presentation time of visual stimuli being .5 sec. Immediately following and in rhythm with the stimulus words, a random 1- or 2-digit number which *S* had to repeat was presented visually. This was necessary because a pilot study had indicated that recall of items in the second set of 4 words was very high. Following the number, the probe was presented in the same modality, and in the same voice and ear or to the same part of the screen as in the original presentation.

Procedure. The *S* read a sheet of typed instructions that described the stimulus presentation. Two sets of instructions were devised, one containing the ST instructions and the other the SP instructions. The ST instructions told *S* to recall the word presented in the same modality and at the same time as the probe word. The SP instructions asked for recall of the word presented in the same place as the probe. It was explained that "same place" referred to the same ear for auditory words and to the same part of the screen for visual words. The *S*s were further instructed that if they could not remember the required word, they were to report any other word from that trial. The *S* was given 16 practice trials, any questions about procedure were answered, and the 64 experimental trials then followed.

The same procedure was followed for the second session. When *S* entered the laboratory, he was given whichever instruction sheet he had not received the first day and instructed to read it carefully. He then received the 16 practice trials and 64 experimental trials, as on the first day.

Results

The following factors were included in the analysis of variance: serial position, modality, instructions, *S*s (nested under the following 2 factors), order of instruction conditions, and order of stimulus material. The dependent variable was the arc sine transformed proportion of words recalled out of the 8 replications (Murdock & Ogilvie, 1968).

Figure 2 shows the probability of recall as a function of serial position, presentation modality, and instruction condition. Serial position refers to the position of the recalled word, not to the position of the probe word. For both auditory and visual items, there is a large effect of serial position, with words in the second set being recalled at a higher level. The probabilities of recall of auditory words in the first and second serial positions were .473 and .617, respectively. For visual words, the

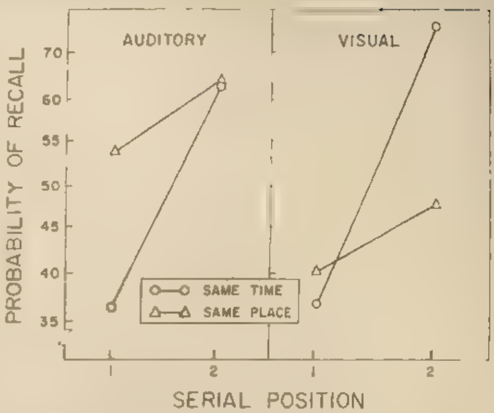


FIGURE 2. Probability of recall as a function of presentation modality (auditory or visual), serial position, and probe type.

corresponding probabilities were .391 and .586. The main effect of serial position was significant, F (1, 3) = 13.6, $p < .05$, as was the Serial Position × Modality interaction, F (1, 3) = 11.0, $p < .05$. The effect of serial position was slightly larger for visual items than for auditory.

The main effect of the instruction condition was not significant overall. However, the instructions did interact significantly with other variables. As Figure 2 indicates, the serial-position effects were quite different for the 2 instruction conditions. This Serial Position × Instructions interaction was confirmed by the analysis of variance, F (1, 3) = 12.6, $p < .05$. When SP instructions were given, the level of recall was only slightly better for the second serial position than for the first. With ST instructions, the difference between recall of words in the first and second serial positions was much larger.

The most interesting aspect of these data is the interaction between presentation modality and instructions. Compare recall of auditory and visual words under the ST instructions. For the first serial position, recall is about equal for auditory and visual items. For the second position, recall is slightly higher for visual items. The SP instructions result in quite a different picture. For both serial positions, recall of auditory items is clearly superior to that of the visual stimuli. The analysis

of variance indicated that the Modality × Instructions interaction was significant, $F(1, 3) = 18.7$, $p < .025$.

The direction of the Modality × Instructions interaction is consistent with the hypothesis of strong associations between successive and auditory stimuli. The forward and backward associations between auditory stimuli are clearly stronger than the corresponding associations for the visual stimuli. These sequential auditory associations are so strong that, for the first serial position at least, recall cued by a backward probe is actually better than that when a simultaneous stimulus is used as a retrieval cue (Duncan's range test, $p < .01$; Edwards, 1968).

That there is no difference in auditory recall for the 2 instruction conditions when the second serial position is tested does not weaken the argument. In the ST condition the item given as a probe was a recent item, whereas in the SP condition the probe was an earlier item. If there were retrieval difficulties involved in probe recognition, these should be greater for the SP probe. Recall of the tested item would then be slightly lower than it would be if the SP probe were as available as the ST probe.

The critical comparison, however, is between auditory and visual presentation under the SP instructions. The large difference found provides evidence for the efficacy of the sequential associations in auditory short-term memory. The difference in auditory recall under the 2 instruction conditions suggests that these sequential associations may, under certain conditions, be stronger than associations based on simultaneity.

The effect of the instruction conditions upon auditory recall becomes even more surprising when possible rehearsal strategies are considered. When the first 4 words were presented, S would probably attempt a quick rehearsal of all 4. Upon presentation of the second set, S would also rehearse those. If an ST probe is given, S only has to retrieve the word he had rehearsed either immediately before or after the probe. However, if S knows he is to get an SP prob[illegible] quandary. He cannot wait until al[illegible] items are presented and then attemp[illegible] rehearse the appropriate pairs, as h[illegible] will probably forget many items b[illegible] rehearses them. Also, the numbe[illegible] be shadowed appears quickly and d[illegible] allow much time for rehearsal. Th[illegible] strategy is to rehearse each se[illegible] it is presented and to then att[illegible] output to recode into the appropria[illegible] rs. Regardless of which strategy S [illegible], the SP condition should give su[illegible] call, as it does with the visual st[illegible] The fact that the SP condition is b[illegible] or auditory stimuli is evidence for th[illegible] ngth of the sequential associations.

Errors. The Ss were in[illegible] to report 1 word on every trial, a[illegible] give any word from the list if th[illegible] not know the correct word. The e[illegible] committed were classified into 4 categ[illegible] (a) same-place error—S responded [illegible] the word that had been presented i[illegible] same place as the probe when he should have given the word that occurred at the same time; (*b*) same-time error—S should have given the word that had occurred in the same place as the probe, but gave instead the word that had occurred at the same time; (*c*) same-modality error—S responded with a word in the correct modal[illegible] but which had been presented neither in the same place nor at the same time as the probe word; and (*d*) miscellaneous—omissions and all other errors. The proportion of each kind of error was tabulated, and the results are shown in Table 1. Proportions are reported rather than frequencies because of the different number of errors in the different conditions. When recall of auditory items was tested, there was a larger proportion of SP errors when an ST probe was given than of ST errors when an SP probe was given. The number of same-modality errors did not differ as a function of instructions. When a visual word was tested, there was a slightly greater proportion of ST errors. There was also a larger proportion of same-modality errors with the ST than with SP instructions.

The data on the types of errors made are consistent with the accuracy data in that they provide evidence for the hypothesized sequential associations in auditory short-term memory. If *S* knew he could not recall the correct word, one would expect him to report at random any word he could remember. Overall, the proportion of ST and SP errors should be approximately the same. The greater number of SP than ST errors and the high frequency of the same-modality errors suggest that it is somehow easier to retrieve an item connected to the probe by means of a forward or backward association than an item that occurred at the same time as the probe.

The occurrence of the sequential associations in auditory short-term memory cannot be regarded as resulting solely from a conscious decision on the part of *S* to process the stimuli in this way. In the present experiment, *S*s knew well in advance which type of probe they would get. They served in only 1 condition per day and had several practice trials before the experiment started. This should have enabled them to encode the stimuli appropriately. In spite of their foreknowledge of the probe condition and the rehearsal opportunities in this task, the SP probe was more effective for at least the first serial position. This suggests that in the time allowed to them, *S*s found it difficult to encode the 2 simultaneous auditory items as a pair, and the sequential associations predominated.

TABLE I

Proportion of Different Types of Errors in Experiment II

Type of error	Presentation modality: Auditory	Presentation modality: Visual
Same time instructions		
Same place	.304	.281
Same modality	.467	.302
Miscellaneous	.229	.417
Same place instructions		
Same time	.235	.340
Same modality	.451	.262
Miscellaneous	.314	.398

Discussion

The results of the present experiments, which used a bisensory–dichotic presentation, extend the findings of earlier studies that employed either the bisensory or dichotic presentation alone or a visual presentation of 2 simultaneous items. The *S*s clearly preferred to recall simultaneous or successive auditory and visual stimuli separately by modality. Two simultaneous visual stimuli are recalled as a pair, but sequences of dichotic auditory stimuli are recalled as 2 sequences, not in pairs. The sequential recall of dichotic stimuli occurs even when visual stimuli intervene between successive auditory stimuli. When a probe technique is used, the SP condition, which tested memory of a forward or backward association, was found to give higher recall for auditory than for visual presentation. For the ST condition, the difference between auditory and visual presentation was smaller and in the reverse direction.

The results of the 2 experiments reported also have implications for a theory of auditory and visual short-term memory. Recently, a number of investigators have hypothesized the existence of 2 separate short-term memory stores, one for auditory items and another containing information in verbal format (e.g., Crowder & Morton, 1969). An alternative form of the hypothesis is that the 2 stores are modality specific, one containing information about auditory items and the other containing visual information (e.g., Murdock & Walker, 1969).

The present data do not support one of these alternative hypotheses over the other. They do, however, suggest an explanation of one of the more puzzling findings in the area of auditory–visual differences. In a serial-probe experiment with lists containing both auditory and visual items, Murdock (1967) found that the best performance was obtained when both the probe and the TBR word were auditory. When either or both were visual, recall was considerably lower. The results of the present experiments indicate that successive auditory stimuli are rather strongly associated, and that this occurs whether or not *S* tries to encode in this

manner. When either or both of the probe and TBR word are visual, this strong association is absent and recall is lower. Thus, the hypothesis of sequential associations between successive auditory stimuli, in addition to accounting for the dichotic and bisensory split-span data, would appear to be useful in integrating within a unified theoretical framework the results of experiments on auditory and visual presentation in short-term memory and those on selective attention.

REFERENCES

BROADBENT, D. E. The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 1954, **47**, 191–196.

BROADBENT, D. E. Successive responses to simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 1956, **8**, 145–152.

BROADBENT, D. E. *Perception and communication*. New York: Pergamon Press, 1958.

BROADBENT, D. E., & GREGORY, M. On the recall of stimuli presented alternately to two sense organs. *Quarterly Journal of Experimental Psychology*, 1961, **13**, 103–109.

BROADBENT, D. E., & GREGORY, M. Some confirmatory results on age differences in memory for simultaneous stimuli. *British Journal of Psychology*, 1965, **56**, 77–80.

BRYDEN, M. P. The manipulation of strategies of report in dichotic listening. *Canadian Journal of Psychology*, 1964, **18**, 126–138.

CROWDER, R. G., & MORTON, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, **5**, 365–373.

EDWARDS, A. L. *Experimental design in psychological research*. (3rd ed.) New York: Holt, Rinehart & Winston, 1968.

GRAY, J. A., & WEDDERBURN, A. A. I. Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 180–189.

MADSEN, M. C., ROLLINS, H. ... Variables affecting imme... sensory stimuli: Ear–eye ... dichotic listening. *Jour... chology Monograph*, 1970, 83 ...

MORAY, N. Broadbent's filt... and the problem of swit... *Journal of Experimental* ... 214–220.

MURDOCK, B. B. Auditory ... short-term memory. *Acta P*... 316–324.

MURDOCK, B. B., & OGILVIE ... variability in short-term mem... *Bulletin*, 1968, **70**, 256–260.

MURDOCK, B. B., & WALKER, ... effects in free recall. *Journa... and Verbal Behavior*, 1969, **8**, 6...

PENNEY, C. G. Bisensory stu... memory: New light on the eff... modality. Unpublished doc... University of Toronto, Canada ...

ROLLINS, H., EVERSON, M., & ... Preferred recall order and re... two messages presented simul... single auditory channel. *Per... physics*, 1972, **11**, 153–155.

SAMPSON, H., & SPONG, P. Bino... immediate memory. *British* ... *chology*, 1961, **52**, 239–248.

SAVIN, H. B. On the successi... simultaneous stimuli. *Percep... physics*, 1967, **2**, 479–482.

SCHURMAN, D. L., EVERSON, M. ... H. A. Successive vs. simultan... of superimposed visual stimuli ... *Psychophysics*, 1972, **11**, 420–422.

WHITMORE, M. G. Stimulus meaning as a variable in dichotic listening. *Psychonomic Science*, 1972, **26**, 207–208.

YNTEMA, D. B., & TRASK, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 65–74.

(Received May 5, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 151-158

ASSOCIATION, DIRECTIONALITY, AND STIMULUS ENCODING[1]

ROBERT E. WARREN[2]

Columbia University

The degree of interference with color naming produced by the word used as a carrier for the color in a discrete-trial Stroop task was found to increase if an associatively related word was concurrently held in memory. The degree of interference varied directly with the associative strength of the connection between the memory word and the color word. A second experiment using unidirectional associates showed that interference with color naming increased only if the word being held in memory elicited the color word and not vice versa. It is suggested that the results support the hypothesis that activation of the memory representations of a word's associates is part of its encoding process.

Attempts to determine the structure of memory have recently centered on the process of stimulus encoding. Encoding as a general term includes all those transformations performed on the nominal stimulus before it assumes its final representation in memory. These transformations may theoretically include both additions to and loss of stimulus information. Encoding is of interest to the memory theorist because of its presumed relation to the structural properties of memory. It has been argued (Wickens, 1970) that the transformations employed in encoding are a function of the structural characteristics of memory. Thus, if one could specify what factors are involved in the initial encoding of words, this information would allow the inference of verbal memory structure.

Various models of memory have suggested different conceptions of the encoding process. Bower (1967) has proposed that encoding involves the specification of values along semantic dimensions and that words are represented in memory as groups of these values. Similarly, Conrad (1967) and others (e.g., Yntema & Trask, 1963) have suggested that encoding involves the amalgamation of particular components or the tagging of a central word unit to indicate specific characteristics. Operationally, however, the nature of encoding has often been assessed in terms of *S*'s reaction to some characteristic of "words as wholes." The proactive interference paradigm used by Wickens (1970), for example, appears to rely on binary decisions regarding the list membership of individual items (Loess & Waugh, 1967). Similarly, false recognition studies (Underwood, 1965) utilize the identification of individual words as old or new as an indication of previous encoding patterns.

The use of whole-word measures is often based on the theoretical position that specification of a particular component or dimensional value affects the representations in memory of not only the presented stimulus, but also of those words sharing the characteristic in question. In the false recognition paradigm, for example, presentation of the word ICE is said to produce implicit activation of words sharing a commonsense impression such as COLD (Kimble, 1968). This position suggests that one way to map out the extent and degree of encoding is to present a word and then test the activation of other words which bear some nominal relation to the presented item. If the words are found to be activated, the common nominal relation could be assumed to be part of the encoding process.

Changes in performance on a number of tasks as a function of word activation

[1] This research was supported by U.S. Office of Education Grant OEG-0-72-0717, U.S. Department of Health, Education, and Welfare, to the University of Oregon, and a Columbia University grant. The opinions expressed herein do not, however, necessarily reflect the position or policy of, or endorsement by, the U.S. Office of Education.

[2] Requests for reprints should be sent to Robert E. Warren, Department of Psychology, Columbia University, 314 Schermerhorn Hall, New York, New York 10027.

can be predicted from a system of word recognition units or logogens proposed by Morton (1970). The logogen system as a whole combines sensory, contextual, and semantic information to identify which word has been encountered in any given situation. The system relies on an initial analysis of sensory information into components which are then routed to logogens which have been activated by similar information in the past. One logogen is assumed to exist for each word in a person's vocabulary, and each is composed of the following 3 elements: (*a*) a counter that increases in value each time stimulus information that corresponds to the word the logogen represents is encountered, (*b*) a motor program that represents the verbal production of the word, and (*c*) a threshold value that is fixed by the past frequency of occurrence of the word. Upon presentation of a word, counter values at logogens which correspond in whole or part to the physical representation of the word begin to increase as components are identified by the sensory analysis system. Counters at those logogens which correspond best to the actual word given increase in value at the greatest rate. When one of the logogens' counter values exceeds its threshold, its motor program is said to become available and is loaded into a response buffer by a central processing unit where it is held for output. Only one program can pass to the buffer at a time, so that if one program becomes available before another, the second will be delayed until the buffer is cleared.

Presentation of a word, then, produces what might be termed primary activation of a logogen. Besides reacting to sensory input, logogens can theoretically also be activated indirectly, as suggested earlier, when words sharing some common semantic feature or other relation are presented. This secondary activation is produced by what Morton (1970) termed the cognitive system. Secondary activation produces counter increments at related logogens, but these increments are thought to be of lesser magnitude and to begin longer after word presentation than those produced by primary activation at th[illegible]n for the word actually encounter[illegible]

In the present experin[illegible]erference with color naming in a di[illegible]l Stroop (1938) task was used to in[illegible]he degree of activation of the wo[illegible] tested In the Stroop task, *S* is [illegible] to name the color of ink in which a[illegible]s printed while attempting to ignor[illegible]em itself In the traditional Stroop [illegible] the base items printed in colors ar[illegible] of Xs or antagonistic color words [illegible] is, the word RED would appear i[illegible]ink, the word YELLOW in red ink, a[illegible]on. The typical finding is that it [illegible]s longer to name the colors when [illegible]se items are antagonistic color nam[illegible] when they are rows of Xs (Jense[illegible] Rohwer, 1966). In an extension of S[illegible]'s work, Klein (1964) has shown tha[illegible] naming latency is also increased ove[illegible] Xs control if common words unrel[illegible] to color are used as base items in the [illegible]

For the logogen system, in[illegible]erence in the Stroop situation occurs because stimulus information activating 2 logogens is present, i.e., those of the base item and the ink color name. If the motor program for the base item becomes availabl[illegible]efore the program for the ink color name, [illegible] production of the color name will be delayed because of the response-buffer space limitations mentioned earlier. The greater the level of activation of the logogen for the base word prior to stimulus presentation, the more likely it is that this will occur. This means that any operation that increases the base word's level of activation prior to its presentation will produce an increase in color-naming latency. It follows that if secondary activation occurs during the encoding of a word, the degree and pattern of activation can be assessed by presenting other words as base items in a Stroop task subsequent to presentation of the word in question and observing whether or not color-naming delays are produced.

Previous work using this technique (Warren, 1972) has shown that the name of the taxonomic category from which a triad of words is drawn, as well as any of the words themselves, produces an in-

crease in color-naming latency when used as the base word in a Stroop task given subsequent to the presentation of the triad. Experiment I tested whether words which appear as responses in association norms produce similar increases in color-naming latencies when used as color base words after their associative stimuli have been encountered by *S*.

To review the argument presented so far, color-naming latency in the Stroop task is thought to vary with the degree of activation of the logogen of the word used as a carrier for the color. If the level of activation of Word B's logogen is somehow increased during the presentation of Word A, the degree of interference with color naming produced when Word B is used as a base item in a subsequent Stroop task should also increase. Experiment I tested whether such an increase in activation occurs when Word A and Word B are related associatively. That is, the following 2 questions are posed: (*a*) Does the associative relationship indicated by standard association norms parallel an underlying pattern of activation in the logogen system such that when a stimulus, Word A, is encountered the logogen of its primary response, Word B, is activated? (*b*) Does the frequency of a response, Word B, to a stimulus, Word A, in an association test reflect the degree or probability of activation of the logogen for Word B when Word A is given?

Experiment I

Method

Subjects. Five male and 5 female students at the University of Oregon were run individually for a single 45-min. session. They were paid $1.50 for their participation.

Procedure. The *S*s were seated at a 2-channel tachistoscope used to present the visual stimuli in the experiment. Prerecorded auditory stimuli were presented to *S*s over earphones.

Each trial in the experiment consisted of 3 phases. In order of occurrence, these were (*a*) auditory word presentation, (*b*) visual word presentation and color naming, and (*c*) recall of the auditory word. In the first phase, the word READY was presented over the earphones followed by a prerecorded auditory word.

Immediately after the presentation of the auditory word, a word printed in colored ink and centered on a white field was shown in the tachistoscope. The word remained visible until *S* had named aloud the color of ink in which it was printed. Naming the color activated a voice-operated relay which terminated the visual display and stopped a Hunter millisecond clock that had been started at the onset of the visual word. The *S*s were instructed to name the color of ink in which a word was printed as rapidly as possible and to ignore the word itself.

In the final phase, which followed color naming directly, *S* recalled aloud the word given auditorily in the first phase.

Both color-naming latency in milliseconds from the onset of the visual item and recall accuracy were recorded. There was a 15-sec. interval between trial onsets.

Design and materials. Thirty pairs of associated words were selected from Palermo and Jenkins (1964). The word pairs and their associative strengths appear in Table 1. Three groups of 10 pairs each were formed with mean associative strengths of .630 (high associative strength), .334 (medium associative strength), and .152 (low associative strength). Associative pairs were chosen so that the range and average frequency of occurrence in English (Kučera & Francis, 1967) of the response members of the 3 groups were approximately equal.

During the experiment, the logogen activation of the response words was tested by presenting them in the color-naming phase of a trial. Interference with color naming produced by each response word was assessed under 2 conditions: (*a*) after the stimulus to which it was a primary associative response had been given auditorily (test condition) and (*b*) after auditory presentation of an unrelated word (control condition). For example, the response word GIRL was preceded in the test condition by the auditory word BOY, while in the control condition it was preceded by the auditory word CASE. The unrelated words used as auditory stimuli in the control condition are shown in Table 1 matched with the assigned response word.

The experiment thus consisted of 6 conditions, resulting from test and control assessment of the color-naming interference produced by each of the response words in the high, medium, and low associative strength groups.

Auditory stimuli in the experiment were prerecorded and presented over earphones. Words presented visually were printed in colored ink in lowercase letters and subtended an average visual angle of less than 3°. The 5 colors of ink used were red, blue, yellow, green, and purple. Five decks of stimulus cards were made up. In each deck, one-fifth of the responses in each associative strength group was printed in each color. Across the 5 decks, each response word appeared in each color. Within a single deck, a response word was

TABLE 1

Auditory and Visual Stimuli: Experiment 1

Group	Auditory stimulus	Visual stimulus	Associative strength	Auditory control
High	LAMP	LIGHT	.706	ADD
	BOY	GIRL	.704	CASE
	SCISSORS	CUT	.678	BOOK
	KING	QUEEN	.651	FOOD
	SLOW	FAST	.634	ART
	BLOSSOM	FLOWERS	.629	HOT
	BED	SLEEP	.577	TIGER
	EAGLE	BIRD	.576	COFFEE
	BUTTER	BREAD	.575	ACTIVE
	HIGH	LOW	.568	QUESTION
Medium	TOBACCO	SMOKE	.482	DOUBLE
	CHAIR	TABLE	.428	HAPPY
	DOORS	WINDOWS	.358	JACKET
	BATH	CLEAN	.354	LAW
	BIBLE	GOD	.316	SHORT
	SQUARE	ROUND	.315	HOME
	SOFT	HARD	.304	RICH
	CRY	BABY	.284	THREAD
	SPEAK	TALK	.252	ARMY
	SMOOTH	ROUGH	.243	LOST
Low	CITY	TOWN	.232	BREAK
	PRIEST	CHURCH	.225	BACK
	GUNS	SHOOT	.201	YOUNG
	WISH	DREAM	.149	HAND
	EARTH	DIRT	.143	LECTURE
	PEOPLE	CROWD	.139	SWIM
	STREET	ROAD	.118	SEAT
	SALTY	SWEET	.107	STONE
	WHISTLE	TRAIN	.106	SKY
	CARRY	HOLD	.102	APPLE

presented in the same color on both its test and control condition appearances. Each deck was used with one-fifth of the *Ss*.

Trials were ordered and presented in 2 35-trial blocks, Blocks A and B, with a 2-min. break between blocks. Each block began with 5 filler trials followed by 30 experimental trials. Within each consecutive set of 6 experimental trials, 1 trial representing each condition was given. Colors were assigned so that no color was repeated on consecutive trials.

All of the response words were shown in each block. Half of the response words in each associative strength group were presented in the test condition and half in the cont[illegible] in Block A. The words were then [illegible]nged and presented in the opposite conditio[illegible] Block B. Half of the *Ss* received Block A [illegible] and half, Block B first. Prior to the begin[illegible] of the experimental trials, *Ss* were show[illegible]mples of each of the 5 colors being used i[illegible] of rows of 5 lowercase Xs printed in colo[illegible]

Results

Recall of the auditory wor[illegible]as good with only 1.5% of the possi[illegible] memory errors occurring (9 out of 60[illegible]ossible). All but 1 of these errors see[illegible] to be acoustic in nature, e.g., recall o[illegible]EAD for THREAD. The 1 remaining err[illegible]was the intrusion of the color word sh[illegible] on the same trial. Because correct [illegible]al perception of the auditory word c[illegible] not be assumed for trials on which m[illegible] errors occurred, these trials were ex[illegible] from further analysis.

Color-naming accuracy was [illegible] good with errors occurring on only 1[illegible] of the trials given (8 out of 600). Co[illegible]naming errors consisted of calling out [illegible] wrong color name (4 8), or prematurel[illegible]ripping the voice key with a hesitation [illegible]nd such as "ah" or "er" (4 8). These t[illegible]ls were also discarded.

Mean color-naming latencies in milliseconds for each of the 6 experimental conditions were determined for each *S*. Overall means are shown in Table 2. A repeated measures analysis of variance was performed on the data with *Ss*, condition (test, control), and associative strength (high, medium, and low) as factors. The analysis showed a significant main effect for condition, $F(1, 9) = 15.44$, $p < .005$, and a significant Condition $\times$ Associative Strength interaction, $F(2, 18) = 6.13$, $p < .01$. The main effect for associative strength did not approach significance, $F(2, 18) = 2.30$, $p > .05$.

Winer's (1962) method for a priori comparisons of treatment totals showed latencies in the high–test condition to be greater than in the high–control condition ($p < .001$). Similarly, medium–test latencies were greater than medium–control latencies ($p < .025$). Latencies in the low–test

TABLE 2

Mean Color-Naming Latencies (in msec.): Experiment 1

Condition	Associative strength group		
	High	Medium	Low
Test	929	906	858
Control	834	856	838

and low–control conditions did not differ significantly ($p > .05$).

In addition to the analysis of variance on within-*S*s means, mean color-naming latencies across *S*s for each stimulus–response pair were determined for both test and control conditions. The ratio of the test to the control mean was then determined for each pair and the correlation between these values and the corresponding associative strengths of the word pairs was calculated. A positive correlation was found, $r = .41$, $p < .05$.

The measure of associative relation used to form the associative strength groups in the experiment was that of forward associative strength, i.e., the frequency with which the auditory word elicted the color word as a response. Backward associative strength was not controlled for, and examination of bidirectional association norms (Keppel & Strand, 1970) shows that a positive correlation existed between the forward and backward associative strengths of the pairs used, $r = .36$, $p < .05$. To examine the influence of this backward associative relation on color-naming latencies, a partial correlation analysis was done on the test–control ratios and the 2 sets of association values for the word pairs in the experiment. The analysis showed the partial correlation of the interference ratios with forward associative strength to be significant, $r = .36$, $p < .05$, while that with backward associative strength was not, $r = .12$, $p > .05$.

Discussion

The findings of Experiment I appear to confirm that secondary activation of words given frequently as responses in association tests does occur during the encoding of the appropriate stimulus word. The degree or probability of activation seems to follow quite closely the frequency with which the word is given as a response. The greater the response frequency, the larger the interference with color naming produced in the test condition.

The word pairs in Experiment I were chosen without regard to the type of associative relation between the words. That is, associations between synonyms (SPEAK–TALK), antonyms (SMOOTH–ROUGH), and words joined by contiguity or function (BED–SLEEP) were all included. It is probably the case that these different associative types reflect different patterns of activation. Differences of this sort have been found in several studies of false recognition (e.g., Grossman & Eagle, 1970). It should also be noted that for all of the pairs of associates used, only primary responses were selected. There is some indication from other studies (e.g., Cramer, 1969) that nonprimary responses do not function in the same manner as primaries.

It has been assumed that it is secondary activation taking place during the presentation of the auditory word in Experiment I that is producing the subsequent interference in the color-naming phase. Since the secondary activation produced by the color word does not become effective until sometime after its presentation, any activation of the recall-word logogen is delayed and unlikely to interfere with color naming. This view is supported by the partial correlation results. Because of the critical nature of the assumption, however, a second experiment which tested it directly was carried out. In Experiment II, word pairs with strong unidirectional associative relations (A→B but B↛A) were selected and the pair member which strongly elicited its mate was assigned to either the recall-word or color-word role in trials similar to those used in Experiment I.

Experiment II

Method

Subjects. Twelve male and 12 female *S*s recruited from the staff and undergraduate populations of Columbia University were run individually for a single ½-hr. session. They were paid $1.75 for their participation.

Procedure. The *S*s were seated in front of a rear projection screen on which all the stimuli in the experiment were shown. A Gerbrands 2-channel projection tachistoscope was used to present the stimulus material.

Each trial in the experiment consisted of 3 phases. In order of occurrence, these were (*a*) recall-word presentation, (*b*) color-word presentation and color naming, and (*c*) recall of the word shown in the first phase.

The word given for later recall was printed in white capital letters on a black background and shown for 800 msec. This exposure was followed by a dark field which lasted 1,000 msec. At the end of this interval, a word printed in colored lowercase letters on a black background was shown and remained visible until *S* had named aloud the

color in which it appeared. Naming the color activated a voice operated relay which terminated the visual display and stopped a Gerbands millisecond clock that had been started at the onset of the colored word. The *S*s were instructed to name the color as quickly as possible and to ignore the word itself. After naming the color, *S* recalled aloud the word shown in white during the first phase of the trial.

Color-naming latency in milliseconds from the onset of the colored word and recall accuracy were recorded. There was a 10-sec. interval between trial onsets.

Design and materials. Two sets of 12 associatively related word pairs were chosen from the bidirectional association norms collected by Keppel and Strand (1970). In both groups, one member of the pair elicited the other as a response, while the opposite relation was weak or nonexistent. For one of these groups (forward association group), the pair member which elicited its mate was assigned to be a recall word during the experimental trials while the other member of the pair was assigned to appear as a color word. For the other group (backward association group), the words were assigned to opposite roles. For the forward association group, the average associative strength for the recall-word–color-word relation was .397 and for the color-word–recall-word relation, .020. For the backward association group, these same values were .021 and .404, respectively. The sets of word pairs are shown in Table 3, along with the individual pair associative strengths. The range and average frequency of occurrence in English (Kučera & Francis, 1967) of the pair members assigned to the color-word role in the 2 groups were approximately equal.

TABLE 3

STIMULI: EXPERIMENT II

Group	Recall word (A)	Color word (B)	Associative strength		Control word
			A→B	B→A	
Forward	KITTENS	CATS	.723	.022	NAIL
	BLOSSOM	FLOWERS	.640	.011	STEAL
	EAGLE	BIRD	.579	.005	COLDER
	DREAM	SLEEP	.485	.082	GRASS
	EASIER	HARDER	.419	.049	WINDOWS
	BATH	CLEAN	.354	.000	SHOOT
	WHISKEY	DRINK	.328	.011	DIRT
	MUTTON	LAMB	.317	.000	WEB
	MOON	STAR	.239	.011	CHAIR
	BROADER	WIDER	.232	.016	NURSE
	STOVE	HOT	.226	.016	THREAD
	PRIEST	CHURCH	.225	.016	PEPPER
Backward	APPLE	FRUIT	.066	.454	TIGER
	FOOT	HAND	.044	.235	SONG
	SWEET	BITTER	.044	.544	HILL
	LAW	JUSTICE	.038	.243	QUESTION
	SOFT	LOUD	.022	.423	SKY
	LIGHT	LAMP	.016	.706	ORDER
	FAST	SWIFT	.005	.452	BABY
	GOD	BIBLE	.005	.316	ALTHOUGH
	HOUSE	COTTAGE	.005	.264	HIGH
	WATER	OCEAN	.005	.362	WAR
	DOWN	SIT	.000	.271	BECAUSE
	SMOKE	TOBACCO	.000	.482	QUEEN

The logogen activation of the color words tested by measuring color-naming latency Experiment I. Color-naming latency for word was determined under the following 2 co tions: (*a*) after presentation of the other men of the associative pair as the word to be rec in the first phase of a trial (test condition) (*b*) after presentation of an unrelated word Phase 1 (control condition). For exampl word STAR from the forward association grou tested after its associate MOON in the test condi and after CHAIR in the control condition. Simil LOUD, from the backward association grou tested after SOFT in the test condition and SKY in the control condition. The unrelated presented as recall words in the control con are shown in Table 3 matched with the appro associative pair.

The experiment, then, consisted of 4 condit resulting from assessment of the color-naming l cies produced by the color words from the forv and backward association groups under test control conditions.

Recall words presented in the first phase trial and the colored words presented in the se phase subtended average visual angles of 2 The 4 colors used for the color words were purple, yellow, and green. Four decks of color- slides were prepared. In each deck, one-fourth the words in each experimental condition appe in each color. Across the 4 decks, each word peared in each color. Within a single deck, a w was presented in the same color on both its test control appearances. Each deck was used for o fourth of the *S*s.

Trials were ordered and presented in 2 28-trial blocks, Blocks A and B, with a 1-min. break between blocks. Each block began with 4 filler trials followed by 24 experimental trials. Within each consecutive set of 4 trials, 1 trial from each condition was given. Colors were assigned so that no color was repeated on consecutive trials.

All of the color words were shown in each block Half of the color words in the forward and backward groups were presented in the test condition and half in the control condition in Block A. The words were presented in the opposite condition in Block B. Half of the *S*s received Block A first and half, Block B first. Prior to the beginning of the first block, *S*s were shown 2 examples of each of the 4 colors being used. Colors were presented in the form of rows of 5 lowercase Xs printed on a black background.

Results

Retention of the recall words was quite good, with only 1.0% of the possible memory errors occurring (11 out of 1,152). The following 4 types of memory errors were noted: (*a*) omissions (4/11), (*b*) intrusion of the color word from the same

TABLE 4

Mean Color-Naming Latencies (in msec.): Experiment II

Condition	Associative direction group	
	Forward	Backward
Test	964	930
Control	892	930

trial (4/11), (*c*) intrusion of a recall word from a previous trial (2/11), and (*d*) intrusion of a color word from a previous trial (1/11). No particular pattern of errors was noted across the 4 conditions. Since correct initial perception of the recall word could not be assumed for trials on which memory errors occurred, these trials were discarded and not used in subsequent analyses.

Color-naming accuracy was also good with errors occurring on 1.7% of the trials given (20 out of 1,152). The following 6 types of color-naming errors were noted: (*a*) naming the wrong color (14/20), (*b*) naming the color word itself (2/20), (*c*) prematurely tripping the voice key with a hesitation sound such as "uh" or "er" (1/20), (*d*) naming the color correctly but too quietly to trip the voice key (1/20), (*e*) calling out the recall word for that trial (1/20), and (*f*) producing no response before the onset of the next trial (1/20). The 2 Type *b* errors occurred on trials in the forward–test condition; the single Type *e* error occurred in the backward–test condition. Trials on which color-naming errors occurred were also discarded.

Mean color-naming latencies in milliseconds for each of the 4 conditions were determined for each *S*. Overall means are shown in Table 4. A repeated measures analysis of variance was done on the data with subjects, condition (test, control), and associative direction (forward, backward) as factors. The analysis showed a significant main effect for condition, F (1, 23) = 8.48, $p < .01$, and a significant Condition × Associative Direction interaction, F (1, 23) = 5.66, $p < .05$. The main effect for associative direction was not significant, F (1, 23) < 1.0, $p > .05$.

Winer's (1962) method for a priori comparisons of treatment totals showed latencies in the forward–test condition to be greater than in the forward–control condition ($p < .001$). The comparison of backward–test and backward–control totals showed no significant difference ($p > .05$).

Discussion

The results of Experiment II appear to support the notion that the color-naming interference that occurs in this paradigm is due to activation of the color base word during the encoding of the recall word. The alternative hypothesis, that the encoding of the color word activates the recall word which then interferes with color naming, does not seem to be supported. Both the presence of the forward–test and forward–control latency difference and the absence of any difference between the backward–test and backward–control latencies contribute to this conclusion.

The lack of any interference in the backward associative conditions does not mean that the logogen for the recall word is not activated after color-word presentation. Rather, it indicates that the time course of activation is such that the color name becomes available and is loaded into the response buffer before the logogen for the recall word passes its availability threshold. Interference in the color-naming task can occur only when a motor program other than that for the color name is loaded in the buffer first. This indicates, as suggested earlier, that secondary or semantic activation does not match in degree or latency the activation caused by sensory input.

Several questions remain to be explored in future research. It is likely that the different types of relations obtained in an association test result from quite different underlying semantic processes. The patterns of activation which parallel these processes may vary both in extent and what might be termed rise time or latency. Experimental variation of semantic relation and of the interval between word presentation and the test for activation may help tease out these encoding patterns. The problem of whether the secondary activation function is set into motion by any amount of logogen activation or whether it only takes place once a logogen passes threshold is also unanswered. There are some indications from other paradigms (e.g., Lewis,

1970) that this latter step may indeed not be necessary.

REFERENCES

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

Conrad, R. Interference or decay over short retention intervals? *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 49–54.

Cramer, P. The determinants of mediated priming: Time interval, semantic and associative relationships. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 266–271.

Grossman, L., & Eagle, M. Synonymity, antonymity, and association in false recognition responses. *Journal of Experimental Psychology*, 1970, **83**, 244–248.

Jensen, A. R., & Rohwer, W. D., Jr. The Stroop color-word test: A review. *Acta Psychologica*, 1966, **25**, 36–93.

Keppel, G., & Strand, B. Z. Free-association responses to the primary responses and other responses from the Palermo-Jenkins norms. In L. Postman & G. Keppel (Eds.), *Norms of word association.* New York: Academic Press, 1970.

Kimble, G. A. Mediating associations. *Journal of Experimental Psychology*, 1968, **76**, 263–266.

Klein, G. S. Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, 1964, **77**, 576–588.

Kučera, H., & Francis, W. N. *Computation analysis of present-day American English.* Providence: Brown University Press, 1967.

Lewis, J. L. Semantic processing of unattend messages using dichotic listening. *Journal Experimental Psychology*, 1970, **85**, 225–228.

Loess, H., & Waugh, N. C. Short-term mem and intertrial interval. *Journal of Verbal Lea ing and Verbal Behavior*, 1967, **6**, 455–460.

Morton, J. A functional model for memory D. A. Norman (Ed.), *Models of human memo* New York: Academic Press, 1970

Palermo, D. S., & Jenkins, J. J. *Word association norms.* Minneapolis: University of Minnesota Press, 1964.

Stroop, J. R. Factors affecting speed verbal reactions. *Psychological Monographs*, **50**(5, Whole No. 225), 38–48.

Underwood, B. J. False recognition produ implicit verbal responses. *Journal of mental Psychology*, 1965, **70**, 122–129.

Warren, R. E. Stimulus encoding and memory. *Journal of Experimental Psychology*, 1972, 94, 90–100.

Wickens, D. D. Encoding categories of w An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

Winer, B. J. *Statistical principals in experimental design.* New York: McGraw-Hill, 1962.

Yntema, D. B., & Trask, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 67–74.

(Received May 14, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 159–160

SHORT REPORTS

PRONOUNCEABILITY AS AN EXPLANATION OF THE DIFFERENCE BETWEEN WORD AND NONSENSE ANAGRAMS[1]

RICHARD S. FINK AND ROGER L. DOMINOWSKI[2]

University of Illinois at Chicago Circle

The present study involved a test of the hypothesis that differences in pronounceability can account for the finding that word anagrams are more difficult to solve than nonsense anagrams. The results, based on 48 anagrams of each type, indicated that word anagrams were more difficult when pronounceability was ignored but that the difference between anagram types was eliminated when pronounceability was statistically controlled. In addition, pronounceability was strongly related to problem difficulty within each anagram class. It was concluded that assessing the ease with which an anagram can be verbally encoded eliminates the need to determine if it has meaning.

An anagram problem involves presenting *S* with a letter sequence and asking him to discover the word which can be made by rearranging the letters. The anagram presented is typically a nonsense string (e.g., given SLACH, find CLASH) but can itself be a word (e.g., given SAUCE, find CAUSE). Previous research has demonstrated that word anagrams are more difficult to solve and that the difference between word and nonsense anagrams can neither be explained in terms of differences in the frequencies of the bigrams constituting the anagrams nor be dismissed as a consequence of methodological biases (Ekstrand & Dominowski, 1968).

In attempting to explain the difference between word and nonsense anagrams, the most obvious approach is to consider the fact that, by definition, word anagrams have denotative meaning and nonsense anagrams do not. However, an alternative, more general, explanation has been proposed. Two studies using only nonsense anagrams have demonstrated that the easier an anagram is to pronounce, the harder it is to solve (Dominowski, 1969; Hebert & Rogers, 1966). The proposal is that, on the average, word anagrams will be easier to pronounce than nonsense anagrams and will be more difficult to solve for that reason. The present study provided a test of this hypothesis.

In order to use a large number of problems, no attempt was made to strictly equate word and nonsense anagrams in terms of pronounceability. The rationale was that, in order to test the hypothesis, all that was needed was reasonably wide variation in pronounceability both among word anagrams and among nonsense anagrams, and overlapping distributions. If the difference between word and nonsense anagrams is to be explained in terms of pronounceability, then it must be true not only that the difference between word and nonsense anagrams is eliminated when pronounceability is (statistically) controlled but also that pronounceability affects problem difficulty within each anagram class.

Method. Using the Olson and Schwartz (1967) tables, 48 word anagrams and 48 nonsense anagrams were constructed. The solutions to the nonsense anagrams were chosen to be similar in language frequency to the solutions of the word anagrams. By using the letter orders of the word anagrams to construct the nonsense anagrams, the number of letter moves required to reach solution (Dominowski, 1966) was equated for the 2 types of anagrams. For each of the 96 anagrams, the mean bigram frequency was calculated, using the Mayzner and Tresselt (1965) norms.

For each problem, 3 additional measures were obtained, using a total of 103 *Ss* all of whom were introductory psychology students fulfilling a course requirement. To allow a check on the comparability of the solutions to the 2 types of problems, the 96 solutions were presented, in random order, to 28 *Ss* who were asked to judge how often each word appears in print, using a magnitude estimation procedure devised by Underwood (1966).

To measure anagram pronounceability, the 96 anagrams were presented one at a time to each of 23 *Ss*, with *S* instructed to pronounce each item as though it were an English word. It was pointed out that some of the items would be common words and some would not, but that *S* was to pronounce every item as though it were a word. Tape recordings of *S*'s responses were made, and the tapes were subsequently run through a Grason-Stadler Model E7300A-1 voice operated relay which triggered Hunter Model A Series D Klockounters. The time taken for each pronunciation was measured in milliseconds.

Problem difficulty was determined by presenting the 96 anagrams to 52 *Ss*, run in 3 small groups. The *Ss* were told what an anagram is, shown examples of word and nonsense anagrams, and told that for each problem their task was to find the word which could be made by rearranging the 5 letters presented. The 96 problems were presented

[1] This article was completed while R. L. Dominowski was a research fellow at the University of Aberdeen, Scotland. The authors are grateful to Sheldon Rosenberg for his help in the measurement of pronunciation times.

[2] Requests for reprints should be sent to Roger L. Dominowski, Psychology Department, University of Illinois at Chicago Circle, Chicago, Illinois 60680.

on a screen in random order. For each problem *Ss* were given 30 sec. to solve and were instructed to mark an X in the answer space if they had not solved a problem at the end of the 30-sec. period.

Results and discussion. The analyses utilized problems ($N = 96$) as the replicate. For each problem, the following measures were available: number of solutions (possible range 0–52), type of anagram (word vs. nonsense), anagram pronunciation time, judged familiarity of the solution, and anagram bigram frequency. Using number of solutions as the criterion, a multiple-regression equation was determined with the other 4 variables serving as predictors. In this analysis, significant contributions were made by only 2 predictors, anagram pronunciation time, $\beta = .64$, $p < .001$, and solution familiarity, $\beta = .16$, $p < .05$. These 2 results confirm previous findings that anagrams that are more difficult to pronounce (= have longer pronunciation times) or that have more familiar solutions are easier to solve (e.g., Dominowski, 1967, 1969; Hebert & Rogers, 1966). Anagram bigram frequency did not independently contribute to the prediction of problem difficulty, a finding consistent with the pattern of earlier results indicating that this variable reflects at best a minor characteristic of anagrams (e.g., Dominowski, 1967).

The distribution of familiarity judgments for word anagrams ($M = 68.6$, $SD = 26.6$) was nearly identical to that for nonsense anagrams ($M = 62.6$, $SD = 25.5$); thus the effect of solution familiarity need not be considered in examining the difference between anagram types. However, as expected, word anagrams ($M = .208$ sec.) were easier to pronounce than nonsense anagrams ($M = .273$ sec.). The results presented above indicate that pronounceability is more important than anagram type (word vs. nonsense) in accounting for anagram difficulty. The following analysis demonstrated this more directly.

If pronounceability was ignored, the mean number of solutions for word anagrams (12.0) was significantly lower than that for nonsense anagrams (16.6), $F(1, 94) = 4.85$, $p < .05$. If, however, pronounceability was used as a covariate, the mean difference was reduced to near zero ($F < 1$). Furthermore, pronounceability was strongly related to problem difficulty both within the class of word anagrams ($r = .61$) and within the class of nonsense anagrams ($r = .62$). These results indicate that pronounceability meets the criteria of a satisfactory explanation of the difference between word and nonsense anagrams. For each type of anagram considered separately, the longer it takes to pronounce an anagram, the easier the anagram is to solve, and the difference in difficulty between word and nonsense anagrams is eliminated when their difference in pronounceability is taken into account.

The results support the conclusion that word anagrams can be considered simply as a set of anagrams which are on the average relatively easy to pronounce. Stated differently, the findin[illegible] gest that there is no need to separate ana[illegible]ms which happen to be words from other anagr[illegible]s of equivalent pronounceability. Although the [illegible]ent results do not support the notion that wor[illegible]nagrams are harder to solve because they hav[illegible]aning, the emphasis on pronounceability is co[illegible]tent with the more general proposal that the mo[illegible]rganized or cohesive an anagram is, the mor[illegible] it will be to solve.

The pronounceability of an anagram [illegible] the extent to which it conforms to the str[illegible] of the language. For 5-letter anag[illegible] sible range is from a string which m[illegible] be considered as 5 units to an anagr[illegible] easily be encoded as a single unit (a [illegible] The accumulated evidence on anagram solv[illegible] that *Ss* first try to think of words based [illegible] rearrangement of the anagram, attemp[illegible] duce syllables when the first meth[illegible] (e.g., Kaplan & Carvellas, 1968). Pres[illegible] [illegible]sily pronounced anagram provides *S* with [illegible] [illegible]perly structured but incorrectly orde[illegible] It is likely that, given such an anagr[illegible] will [illegible]ntain its plausible but incorrect or[illegible] to find the solution. In contrast, [illegible]orly structured anagram (e.g., a string [illegible]ants followed by a vowel), *S* is unlikel[illegible] its order but will rather quickly rearra[illegible]ted letters into a plausible basis for a word [illegible] also possible that *S* will simply perseverate o[illegible] pronounced anagram; i.e., solution att[illegible] will be made at a slower rate, at least initially. [illegible]hough it seems unlikely that this mechanism alo[illegible] is responsible for the effect of pronounceability. [illegible] the appropriate record were obtained, it would b[illegible] possible to determine if anagram pronounceability affects both the rate and the nature of solution attempts.

REFERENCES

Dominowski, R. L. Anagram solving as a function of letter moves. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 107–111.

Dominowski, R. L. Anagram solving as a function of bigram rank and word frequency. *Journal of Experimental Psychology*, 1967, **75**, 299–306.

Dominowski, R. L. The effect of pronunciation practice on anagram difficulty. *Psychonomic Science*, 1969, **16**, 99–100.

Ekstrand, B. R., & Dominowski, R. L. Solving words as anagrams: II. A clarification. *Journal of Experimental Psychology*, 1968, **77**, 552–558.

Hebert, J. A., & Rogers, C. A., Jr. Anagram solution as a function of pronounceability and difficulty. *Psychonomic Science*, 1966, **4**, 359–360.

Kaplan, I. T., & Carvellas, T. Effect of word length on anagram solution time. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 201–206.

Mayzner, M. S., & Tresselt, M. E. Tables of single letter and bigram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, **1**, 13–32.

Olson, R., & Schwartz, R. Single and multiple solution five-letter words. *Psychonomic Monograph Supplements*, 1967, **2**, 105–152.

Underwood, B. J. *Experimental psychology*. (2nd ed.) New York: Appleton-Century-Crofts, 1966.

(Received December 8, 1972)

Journal of Experimental Psychology
1974, Vol 102, No. 1, 161-163

WHEN IS RECALL SPECTACULARLY HIGHER THAN RECOGNITION? [1]

MICHAEL J. WATKINS [2]

Yale University

Tulving has described a paired-associate procedure in which the B terms were recalled better than they were recognized. The present experiment replicated and magnified this finding. It is considered that these results argue against current models in which recall and recognition are characterized as qualitatively distinct processes

There is currently some disagreement about whether recall and recognition are to be accounted for by the same or by different mechanisms. The predominant view postulates 2 qualitatively distinct processes (Kintsch, 1970; Underwood, 1972). According to this theory, recall involves a search-cum-implicit-retrieval stage, followed by a decision stage, in which the retrieved information is either accepted as appropriate or rejected. In recognition, the retrieval of an available item is inevitable (or "automatic"), and hence only the decision phase is functionally operative. Other students of memory reject the notion that retrieval is bypassed in recognition. Thus, Tulving and Thomson (1971) argue that while retrieval is usually comparatively more successful in recognition, the retrieval phase is nonetheless a distinct and fallible component of the recognition process, just as it is in recall.

One implication of the dual-process theory is that recognition should be higher than recall. The only exception is when there is no retrieval failure in recall, in which case the 2 measures should be identical. In a paper entitled "When is recall higher than recognition?" Tulving (1968) reported an experiment in which *S*s learned a paired-associate list to a criterion of 2 successive error-free trials, and were then given a recognition test of the B terms. Approximately 10% of the B terms were not recognized. Tulving's results are clearly difficult to interpret in terms of the dual-process theory. That the recall advantage was obtained with a paired-associate procedure rather than with free recall does not detract from its theoretical importance. To be recalled, an item must be represented or "available" in memory, and an automatic-access theory of recognition predicts recognition of *all* available items. However, Tulving's recall superiority was rather small and it might be held that the experiment was not totally conclusive, in that (*a*) incorrect identification was less frequent than failure to identify a presented item, and hence the recall advantage would have been even less if *S*s were forced to make more recognition responses; and (*b*) a few items may have become unavailable during the course of the recognition test, perhaps as a result of considering the distractor items.

[1] This research was supported by National Science Foundation Grant GB-24171X. I am indebted to my wife Olga for collecting and analyzing the data, and to Endel Tulving for encouraging the publication of the paper.

[2] Inquiries concerning this article should be sent to Michael J. Watkins, Department of Psychology, 333 Cedar Street, Yale University, New Haven, Connecticut 06510.

In the demonstration to be reported, the procedure was free from these potential sources of artifact, and yet yielded a greatly magnified recall advantage.

Method. Twenty male and female undergraduate psychology students were tested individually or in small groups. In outline, all *S*s were presented with and tested on a practice paired-associate list, and were then shown an experimental list of which certain pairs had been designated critical. Immediate testing was confined to noncritical pairs. After a filled interval all *S*s were given free and forced choice recognition tests of the critical B terms. Finally the critical A terms were presented in a conventional cued recall test of the corresponding B terms.

For both the practice and experimental lists, the A terms contained 5 letters and the B terms 2 letters. The majority of both the A and B terms were "nonsense words," insofar as they do not appear in the dictionary. However, if each B term was added to its A term, a common word was formed. This was true for all pairs of both practice and experimental lists. Neither A nor B terms were repeated within or between lists. There were 20 pairs in the practice list, while the experimental list contained 39 pairs of which 24 were designated critical and 15 noncritical. The *S*s were randomly assigned to 2 groups of 10. Both groups saw the same practice list but different forms of the experimental list. The 2 variants of the experimental list differed only in the B terms of the critical pairs. Examples of critical pairs presented to one group of *S*s are SPANI - EL, EXPLO - RE, LIQUE - FY, HIDEO - US and AMNES - IC; the corresponding pairs for the other group would be SPANI - SH, EXPLO - DE, LIQUE - UR, HIDEO - UT, and AMNES - TY. The *S*s were not explicitly informed of the nature of the relationship between A and B terms, although this was presumably perfectly obvious from the first 1 or 2 pairs of the practice list. The order of the A terms for both critical and noncritical pairs was identical for the 2 groups of *S*s; the first and last 5 pairs were noncritical, while the remaining 5 noncritical pairs were randomly allocated among the 24 critical pairs.

Presentation was by means of a closed-circuit television system, with successive pairs typed in uppercase, the A term on the left, the B term on the right, and a hyphen between the 2 terms. The pairs were presented at a 4-sec. rate.

Immediately following the presentation of the practice list, *S*s were given a list of all 20 A terms,

in a new random order. The *Ss* wrote down, at their own rate, as many of the B terms as they could recall.

The experimental list was simply announced as "next list" and the *Ss* were led to believe that they would be tested in just the same way as for the practice list. In fact at this stage they were merely given a recognition test of the 15 noncritical pairs. The test consisted of the presentation of the 15 noncritical pairs plus 15 new pairs of the same type. The *Ss*' task was to circle 15 pairs which they thought most likely to have appeared in the preceding list. The purpose of this test was simply to reduce the probability of covert retrieval and rehearsal of critical pairs, and the impression was given that this test terminated the experiment.

The *Ss* next participated for 30–40 min. in a separate experiment, which involved the free recall of a series of word lists. Following this experiment, each *S* was unexpectedly given a list containing the 24 critical B terms seen in the experimental list, randomly mixed with the 24 critical B terms which had been presented to the other group of *Ss*. Instructions were first given for a free recognition test; *Ss* were to check any item recognized as occurring in the second list. The *Ss* were then given a different pen and asked to check further items which they thought most likely to have been presented before, until a total of 24 had been checked in this forced recognition test. Finally the *Ss* were given a cued recall test of the 24 critical pairs. The 24 A terms were presented in random order, and *Ss* were instructed to write down the B terms with which they had been paired. Thus for both groups of *Ss* the recognition and cued recall tests were identical, although the correct responses differed. In both tests *Ss* were given as much time as they needed.

The confounding of type of test and order of test may have given a small advantage in favor of the recognition test. However, this effect is not critical in that it would merely serve to reduce the predicted recall advantage. The converse possibility that exposure to the recognition items may have enhanced recall seems unlikely, in that the recall targets and the recall lures were equally exposed in the recognition test.

Results. Essentially identical patterns of results were obtained for the 2 groups of *Ss*. The 2 sets of data were therefore pooled. For all 20 *Ss* the number of presented critical B terms successfully recalled was greater than the number freely recognized, and equaled or exceeded the number responded to in the forced choice recognition test despite the fact that half of the items in the latter condition could have been identified by chance. Table 1 shows the mean probability of correct response for both presented and nonpresented (control) items in the recognition and recall tests. The control probabilities were used to estimate the amount of guessing in each of the test conditions. In detail, the difference between the presented and control probabilities for each test condition was divided by the probability of not responding to a control item. For example, the corrected free recognition probability was given by .14 − .07, divided by 1 − .07, that is, .08. These corrected probabilities are also shown in Table 1.

TABLE 1

PERFORMANCE AS A FUNCTION OF MEMORY TEST

Item	Free recognition	Forced recognition	Cued recall
Probability of correct response	.14	.52	.72
Probability of incorrect response	.07	.48	.14
Corrected probability of correct response	.08	.09	.67

Discussion. The results clearly support Tulving's (1968) conclusion that in some circumstances information may be recalled better than it can be recognized. Moreover, the present recall advantage was large. It is fun to speculate how the present paradigm might be refined to further accentuate this advantage—perhaps to perfect recall and zero (chance) recognition. Recall performance could presumably be increased by reducing the retention interval, or else by improving the original learning. One way of enhancing the initial learning of the A–B pairs would be to reduce the presentation rate, while a second would be to instruct the *Ss* to form a visual image comprising the A and B terms in interaction (Bower & Winzenz, 1970)—a strategy to which the present material would lend itself particularly well. Recognition performance could probably be reduced by further encouraging attention during presentation to be focused on the A–B pairs as such, and less on the individual terms. One way of achieving this would be to present to one group of *Ss* A–B pairs such as PLUN*ER - ****G**, REVEN*E - *****U*, etc. while the second group would see PLUN*ER - ****D**, REVEN*E - *****G*, etc. It is likely that the B terms would be readily retrieved on presentation of the A terms, yet poorly recognized.

Of course the technique being exploited with such a procedure, as in the experiment reported, involves making only the A–B *pairs* meaningful, thereby encouraging *Ss* to pay little attention to the individual A and B elements. It is conceded that such contrived procedures are not representative of those used in most verbal memory studies, and that the finding of recall that is better than recognition is likewise unusual. But the finding is not without theoretical significance, in that it poses a difficulty for the dual-process theory. According to this theory, failure to recognize an item indicates that it is not available in memory; however, in the present experiment the availability of a substantial proportion of nonrecognized items was demonstrated by their subsequent recall. Moreover, to interpret the obtained recall advantage in terms of the organization of nominal or *E*-defined units into very different functional units is of little comfort to the dual-process theorist, for "If retrieval and recognition processes can be separated as neatly as the dual process theory implies, organization of the

learning material can have no effect upon recognition, since organization facilitates retrieval and retrieval in recognition is trivial [Kintsch, 1970, p. 280]."

It might be thought that the present results simply demonstrate a tendency to encode material into whole words rather than partial ("nonsense") words. While this may be largely true for the present experiment, it is argued here that hierarchical structuring does not terminate at the level of the word; rather the general principle of organizing or chunking is emphasized. In fact, in the present experiment 20% of the B terms were words in their own right, yet were apparently not well encoded as separate elements. Moreover, in Tulving's (1968) experiment all the terms (both As and Bs) were meaningful words, yet at least some were encoded in a higher order unit.

One way of conceptualizing the present findings is to assume that an event is remembered only when a retrieval cue makes contact with the stored representation of the event. With paired-associate and recognition procedures, the retrieval cue is the perception of the A term and nominal "target," respectively. In free recall, the nature of the retrieval cue may not be obvious. In all cases, however, successful retrieval follows only when the retrieval cue matches the item trace, as originally encoded (Tulving & Watkins, in press). According to this view, when is recall higher than recognition? The answer is ". . . whenever retrieval cues present at the recall test are more effective in providing access to stored information than are retrieval cues present at the recognition test [Tulving, 1968, p. 54]." And when is recall spectacularly higher than recognition? Whenever the difference in effectiveness of the 2 cues is a spectacularly large one.

REFERENCES

Bower, G. H., & Winzenz, D. Comparison of associative learning strategies. *Psychonomic Science*, 1970, **20**, 119-120.
Kintsch, W. *Learning, memory and conceptual processes.* New York: Wiley, 1970.
Tulving, E. When is recall higher than recognition? *Psychonomic Science*, 1968, **10**, 53–54.
Tulving, E., & Thomson, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116 124.
Tulving, E., & Watkins, M. J. Continuity between recall and recognition. *American Journal of Psychology*, in press.
Underwood, B. J. Are we overloading memory? In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, D.C.: Winston, 1972.

(Received December 29, 1972)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 164–166

APPARENT SLOWING OF BIMANUALLY ALTERNATING PULSE TRAINS[1]

SEYMOUR AXELROD[2] AND MICHAEL NAKAO

Department of Psychiatry, State University of New York at Buffalo

Auditory clicks presented in rapid alternation to the 2 ears are heard as slower than clicks presented at the same rate to a single ear. We examined here a similar phenomenon in somesthesis. Thirty *S*s compared the rates of bimanually alternating taps with the rates of taps presented to a single hand, or vice versa. In terms of inter-*S* consistency, of the dependence of degree of perceptual slowing on repetition rate, and of the quantitative aspects of the dependence, the somesthetic results were similar to those obtained in an earlier auditory experiment, suggesting that the apparent slowing in the 2 modalities of stimuli alternating between receptive surfaces on the 2 sides of the body has a common basis central to modality-specific afferent structures.

When rapidly repeated clicks are presented alternately to the 2 ears, their apparent repetition rate is slower than that of clicks presented at the same objective rate to one ear alone (Axelrod, Guzy, & Diamond, 1968). Similarly, the numerosity of dichotically alternating stimuli is less accurately reported than that of monotic ones (Guzy & Axelrod, 1972; Holden, 1973).

These phenomena reveal a limitation in the processing of bilaterally alternating stimuli. If such limitation is unique to audition, then it would make sense to suppose that the underlying neural mechanisms are located in audition-specific structures, and that the limitation thus arises relatively early in the series of events from stimulus to perception. Explanations might then invoke the sort of contralaterally inhibitory process found, e.g., in single units of the cat's inferior colliculus (Erulkar, 1959). But if a similar limitation is demonstrable in another modality, then it would be possible (though not necessary) to posit a common limiting mechanism, located central to modality-specific pathways, and operating later in the processing sequence.

Method. Subjects were 30 students, 16–30 yr. old, whose participation fulfilled requirements of courses in introductory psychology. None admitted to any somesthetic deficit.

Stimuli were impacts delivered by headless 6/32 screws inserted into the driving spindles of 2 Ling (V 47, Model 102) vibration generators. Each generator was covered by a cap with a central hole of 8-mm. diameter; when *S* lightly rested the palmar surface of the middle fingertip on the hole, the finger made light contact with the screw. Two-millisecond square pulses gated a transistor which impressed 6 v. across the generator and a series resistance. There were 2 such circuits, one for each hand, the series resistances being adjusted to make the impacts on the 2 hands subjectively equal in intensity to *E*s. The distance between the screws delivering the impacts was 15 cm. The sounds made by the generators were masked by a low level of thermal noise.

Half the *S*s (9 male, 6 female) constituted the unimanual standard (US) group, and half (8 male, 7 female) the bimanual standard (BS) group. The *S*s in the US group were told that they would feel 2 4-sec. trains of taps, a standard train to the left hand, followed after a .5-sec. interval by a comparison train in which the taps alternated between the 2 hands, starting with the left. The task was to judge whether the comparison taps occurred at a rate faster than, equal to, or slower than that of the standard. A method of limits was used, the midpoint of the interval of uncertainty being taken as the point of subjective equality (PSE). The task was demonstrated in a trial using a long (2,000-msec.) standard interstimulus interval (ISI).

Three onset-to-onset ISIs were used as standards: 1,000, 250, and 100 msec. At each ISI, 5 ascending trials alternated with 5 descending trials; within a trial, the comparison ISI was varied in equal steps, each 10% of the standard. Sequencing of trials was such that effects of practice, fatigue, etc., were distributed over ISI, within and between *S*s. The conditions for the BS group were essentially the same as for the US group, except that these *S*s had to judge a unimanual comparison train relative to a bimanual standard.

The experiment was thus similar to the auditory experiment of Axelrod et al. (1968), the major differences being that in the earlier study, (*a*) a wider range of ISI was explored (1,000–25 msec.), and (*b*) the comparison ISIs in a trial varied in a logarithmic series. The linear series used here provided a denser set of comparison ISIs around the standard than that used in the auditory experiment.

Results. For each *S* for each ISI, the score $[(\bar{X}_S - \text{standard}) \times 100/\text{standard}]$ was computed, where $\bar{X}_S$ is the mean of *S*'s 10 PSEs. The score represents the percentage deviation of *S*'s mean PSE from the standard ISI. The means of these scores are given in Table 1. All 6 means are in the

[1] This study was supported by Research Grant 06862 from the National Institute of Neurological Diseases and Stroke, and by General Research Support Grant RR05400 from the General Research Support Branch, Division of Research Resources, National Institutes of Health. Michael Nakao was supported by a fellowship from the United Health Foundation of Western New York. Thanks are due to Donald Harmon for advice on instrumentation, to John Tank and Eric Zinnerstrom for technical assistance, and to Dewey Bayer and Janet Crockwell for assistance in procuring *S*s.

[2] Requests for reprints should be sent to Seymour Axelrod, Department of Psychiatry, SUNY, 2211 Main Street, Buffalo, New York 14214.

directions indicating apparent slowing of bimanually alternating trains relative to unimanual ones, and each is significantly different from the corresponding standard, *ts* (14) $\geq$ 4.04, *ps* < .01. In both groups, mean percentage deviation increased monotonically as standard ISI decreased, i.e., relative slowing increased with repetition rate. With one exception in each group, both with 1,000-msec. standards, all *Ss* showed relative slowing of bimanual pulses at all standard ISIs; by binomial test, the proportion 14/15 has $p < .001$.

A separate Treatments × *Ss* analysis of variance was performed on the percentage deviations for each group. In both analyses, the effects of ISI were significant, *Fs* (2, 28) $\geq$ 11.50, *ps* < .001. For the BS group, the difference between the means at 1,000 msec. and 250 msec. was significant by one-tailed test, t (28) = 1.84, $p < .05$. For the differences between the 250-msec. and the 100-msec. conditions, and between the 1,000-msec. and the 100-msec. conditions, *ts* (28) $\geq$ 2.92, *ps* < .01. For the US group, the difference between the means at 1,000 msec. and 250 msec. was not significant; for the remaining 2 differences, *ts* (28) $\geq$ 4.48, *ps* < .01.

The mean PSEs for each group were transformed from measures of period (ISI) to measures of rate (taps per second), and the resultant group data are plotted in Figure 1. The triangles directly represent the data of the BS group, with the bimanual standards on the abscissa and their corresponding PSEs on the ordinate. For US group (circles), bimanual PSEs are on the abscissa, with the unimanual standards on the ordinate. This is the same procedure as that used in analyzing the auditory data (Axelrod et al., 1968, pp. 52–53). The line of best fit to the 6 points has a y intercept of .38 and a slope constant of .68. The light solid line represents the function obtained with clicks by Axelrod et al. (1968), with intercept = .05 and slope = .62.

Individual slope constants were calculated for the 3 data points for each *S*. The means for the 2 groups were not significantly different; the grand mean was .68, agreeing with the group data value. The difference between the grand mean and the hypothetical veridical slope (1.00) was significant, t (29) = 12.31, $p < .001$, reflecting the uniform direction and significance of the percentage deviations from zero in Table 1. Likewise significant was the difference between the obtained slope and the value (.50) which would have resulted had *Ss* systematically ignored or suppressed all the pulses from one hand during bimanual stimulation, t (29) = 6.92, $p < .001$.

The group means of the individual intercepts were not significantly different. The grand mean was .42, approximating the group data value (.38), and was significantly different from zero, the intercept of the veridical line, t (29) = 6.42, $p < .001$.

To permit comparison of these results with those of the auditory study of Axelrod et al. (1968), we calculated slopes and intercepts for each of the 40 *Ss* in that study, from the original data. The mean slope coincided with the group data value (.62).

TABLE 1

MEAN PERCENTAGE DEVIATIONS AND STANDARD ERRORS FOR ALL CONDITIONS

Standard interstimulus interval (in msec.)	Group: Bimanual standard		Group: Unimanual standard	
	$\bar{X}$	$S_{\bar{X}}$	$\bar{X}$	$S_{\bar{X}}$
1,000	9.87	2.44	−12.67	2.85
250	24.40	5.98	−14.30	2.34
100	47.07	7.30	−28.67	3.24

Note. Each mean is based on 15 scores.

The difference between the somesthetic mean and the auditory one was not significant, t (68) = 1.50, $.10 < p < .20$. The mean intercept in the auditory study was .21 (group data value .05), and the difference between the means in the 2 modalities was again not significant, t (68) = .86. Therefore, for neither of the 2 determining constants (slope and y intercept) is there reason to assume a difference between the modalities; i.e., the auditory and somesthetic lines derived from the individual data are not significantly different from coincident.

Discussion. The present results were thus strikingly like those obtained in audition. (*a*) Apparent slowing of alternating stimuli was consistent across *Ss*. (*b*) The degree of slowing (percentage deviation) increased monotonically with repetition rate. (*c*) The group data points on the plot of single-source PSE as a function of alternating rate were well fitted by a straight line, (*d*) whose slope approximated that obtained with clicks. While there was a difference of $\frac{1}{3}$ of a pulse per second between the intercepts of the auditory and somesthetic group data lines, (*e*) there were no significant modality differences between the mean slopes or mean intercepts calculated from the individual data.

Although it can be argued that analogous processes might occur in modality-specific portions of the auditory and the somesthetic pathways, the similarity of the findings in the 2 modalities is consonant with a common central limitation on the processing of bilaterally alternating stimuli. The limitation might arise from a relative inefficiency in shifting attention rapidly between spatially separated sources. In speculation on possible bases of such an inefficiency in audition, it has been suggested (Guzy & Axelrod, 1972; Kahneman, 1970) that breakdown in attention shifting may be due to inability of motor orienting responses to follow the stimulus-alternation rate; such a mechanism would operate equivalently in audition or somesthesis, and generate parametrically similar results in the 2 modalities.

An alternative approach comes from the demonstration by Bregman and Campbell (1971) that *Ss*, listening to a repeated sequence of 6 tones (a subset of 3 high-pitch, interdigitated with a subset of 3 low-pitch, tones) presented at 10 per second, split the sequence into 2 subjectively segregated, nonintegrable streams—a high and a low. Auditory

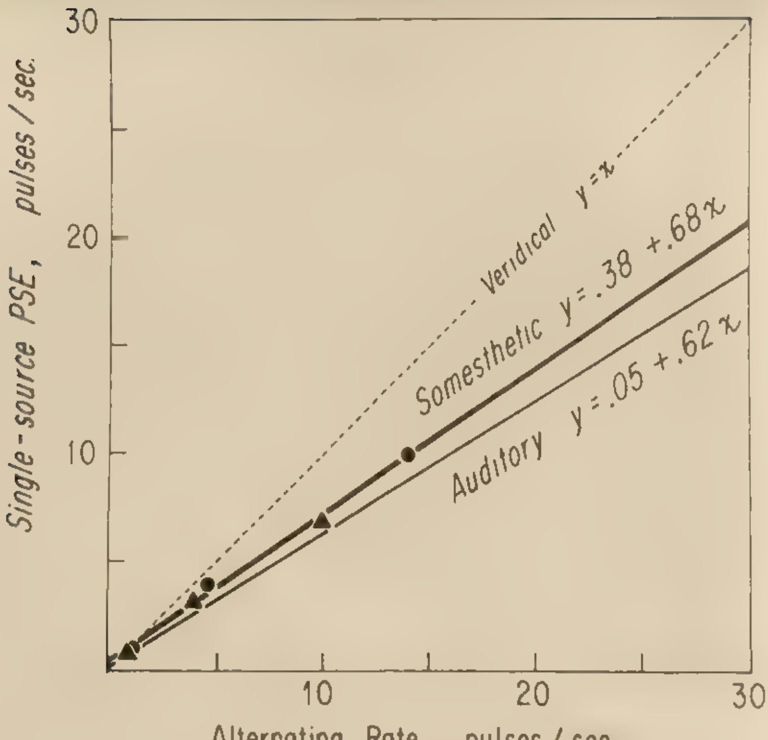


FIGURE 1. Single-source point of subjective equality (PSE) as a function of alternating-source rate. (The heavy line represents unimanual PSE as a function of bimanual rate [group data], and is the line of best fit to the data points. Triangles: bimanual standard group. Circles: unimanual standard group. The light solid line is the line of best fit to the auditory-click data of Axelrod, Guzy, and Diamond, 1968.)

and somesthetic slowing of bilaterally alternating trains may reflect an inability to integrate the results of similar segregations into left and right streams, a possibility which could be tested by using variants of the Bregman–Campbell paradigms with ear or body-side subsets, rather than pitch subsets. Stream segregation by spatial location appears to have been implied in von Békésy's (1959) early observations that when white-noise bursts are presented alternately between widely spaced speakers, or between widely spaced somesthetic stimulators on an arm, the observer can attend to one or the other source, but not simultaneously to both.

So far as neural substrate is concerned, 2 types of account of the present results suggest themselves. One would involve a central mechanism which is less able to integrate bilaterally alternating stimuli than to process single-source stimuli, e.g., because of greater refractoriness to alternating than to unilateral inputs. A mechanism of this sort would presumably be located in an area receiving both auditory and somesthetic information, such as association cortex (Thompson, Johnson, & Hoopes, 1963) or superior colliculus (Wickelgren, 1971). Another possibility is contained in Holden's (1973) suggestion that difficulty in integrating alternating stimuli is a function of the disparateness of the cortical projections of the stimulus subsets. This hypothesis is especially attractive in being testably applicable to many dimensions from which subsets can be drawn —e.g., ear, pitch, location in auditory or visual space, body locus, and sensory modality.

REFERENCES

AXELROD, S., GUZY, L. T., & DIAMOND, I. T. Perceived rate of monotic and dichotically alternating clicks. *Journal of the Acoustical Society of America*, 1968, **43**, 51–55.

BREGMAN, A. S., & CAMPBELL, J. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 1971, **89**, 244–249.

ERULKAR, S. D. The responses of single units of the inferior colliculus of the cat to acoustic stimulation. *Proceedings of the Royal Society of London, Series B*, 1959, **150**, 336–355.

GUZY, L. T., & AXELROD, S. Interaural attention shifting as response. *Journal of Experimental Psychology*, 1972, **95**, 290–294.

HOLDEN, E. A., JR. Effects of stimulus alternation on sequential information processing by retarded and nonretarded subjects. *Journal of Experimental Psychology*, 1973, **101**, 262–268.

KAHNEMAN, D. Remarks on attention control. *Acta Psychologica*, 1970, **33**, 118–131.

THOMPSON, R. F., JOHNSON, R. H., & HOOPES, J. J. Organization of auditory, somatic sensory, and visual projection to association fields of cerebral cortex in the cat. *Journal of Neurophysiology*, 1963, **26**, 343–364.

VON BÉKÉSY, G. Similarities between hearing and skin sensations. *Psychological Review*, 1959, **66**, 1–22.

WICKELGREN, B. G. Superior colliculus: Some receptive field properties of bimodally responsive cells. *Science*, 1971, **173**, 69–72.

(Received January 10, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 167-169

LEARNER JUDGMENT IN INSTRUCTIONAL DECISIONS FOR LEARNING MEANINGFUL PAIRED ASSOCIATES[1]

M. I. CHAS. E. WOODSON[2]

University of California, Berkeley

The problem studied is the facilitation of the learning by college students of meaningful paired associate lists consisting of pairs of romanized Japanese words and corresponding English words. Six instructional strategies differing in the method of determining the frequency of item occurrence were evaluated experimentally. The strategies differed in that during instruction, the next item was (*a*) randomly selected, (*b*) randomly selected with item frequencies balanced to give more training on difficult items, (*c*) randomly selected from among items that had not been correctly responded to twice consecutively, (*d*) learner selected where learners received no instructions about selection, (*e*) learner selected where learners were told to select items they did not know, and (*f*) learner selected where learners were told to select items they knew. Learning, as measured by a retention test 1 wk. later, was superior for the sequences which lead to high error rates during instruction, whether or not they were learner selected or computer selected. Learner's judgment can be used to increase instructional effectiveness during instruction and can be made more effective by instructions to the learner. Implications for instructional theory are discussed.

Atkinson (1972b) has reported an experiment in which a model for computer-selected instructional sequences was found to be more effective than learner-selected or randomly determined instructional sequences. Atkinson found learner selection of instructional sequences considerably less effective than a computer model based on a theory of learning which postulates 3 states of learning and used transition probability matrices, estimated for each item, for state changes when an item is responded to correctly and when another item is missed. One might take the position that his learners were handicapped in their selection by not following the general instructional strategy used by the computer model, that is, that it is most effective to study those items which one does not know very well. The average number of errors during training show *Ss* drilled themselves on relatively learned material about midway between the most difficult and easiest computer-selected sequences, perhaps out of motivation to get a higher score during instruction or perhaps out of the general habit of opting to respond to what one feels one knows, a habit which is probably taught indirectly in school.

The instructional strategies discussed here may be considered schemes for the allocation of instructional trials among items. The basic strategy is to allocate more trials to unlearned items and fewer to learned items. Presumably if the learner is able to select those items which he does not know and those which he knows, his judgment would be very valuable as a basis for the allocation of instructional trials among items.

The purpose of the present study is to examine the question of whether learner-selected instructional sequences can be made more effective if the learners are given specific suggestions about a general strategy for selecting their instructional experiences.

Method. The *Ss* were 120 undergraduate students; 20 *Ss* were assigned randomly to each of 6 instructional conditions and were instructed individually. None of the *Ss* had prior course work in Japanese and none professed familiarity with the language.

The instructional materials were 50 romanized Japanese words paired with English words. The pairs included 30 nouns, 11 verbs, and 9 adjectives, and were selected to have nearly equivalent meanings in the 2 languages. The words involved simple everyday terms (uchi, house; yoru, night), in most cases having 4 or 5 letters. The 50 Japanese words were divided randomly into 5 subsets of 10. The order of words within these 5 subsets was randomly determined for each *S* and these lists were before *S* during the training sessions.

The *Ss* participated in 2 sessions: an instructional session which lasted approximately 2 hr. and a test session, 1 wk. later, which lasted about 20 min. The instruction was conducted by teletypewriters and an HP2000B computer (Hewett-Packard) time-sharing system.

The instructional session involved a series of self-paced trials, each involving a response to a particular item. Each *S* had before him, mounted for easy reading on the teletype terminal, 5 lists of Japanese words which remained in view throughout the training period. The instructions to *S* explained that he must respond to one of the items from each of the lists in sequence from list to list. In the computer-controlled sequences this was automatically con-

[1] This research was supported in part by a grant from the National Institute of Mental Health to the University of California Institute of Human Learning and by computer services provided by the Lawrence Hall of Science, University of California, Berkeley.

[2] Requests for reprints should be sent to M. I. Chas. E. Woodson, Institute of Human Learning, University of California, Berkeley, California 94720.

TABLE 1

Proportion of Correct Responses in Successive Trial Blocks during the Instructional Session and on the Retention–Transfer Test Administered One Week Later

Condition	Successive trial block (instructional session)					Test
	1	2	3	4	5	
R	.15	.41	.61	.65	.70	.40
RD	.07	.17	.27	.24	.32	.80
RU	.05	.10	.14	.13	.19	.79
LS	.11	.20	.27	.32	.36	.52
LSK	.30	.41	.37	.50	.67	.22
LSU	.03	.06	.12	.10	.15	.85

trolled as the computer selected the next Japanese word. In the learner-selected conditions this was not controlled, but a spot check indicated *S*s were highly reliable in following this procedure.

In all instructional conditions *S* responded by typing the Japanese word, a blank, and the English word. If the response by *S* was correct, the system proceeded to the next item. If the response by *S* was incorrect, the system crossed out his answer and typed the correct answer, thereby providing feedback regarding *S*'s responses. The 6 experimental groups were distinguished by the way the next item was selected for instruction:

In the random order (R) condition, the next item was selected randomly (without replacement) from the appropriate sublist of Japanese words.

In the random order balanced for difficulty (RD) condition, the next item was selected randomly with the item distribution balanced to make more difficult items occur more frequently. Data from Condition R were used to estimate the mean trials to learning for each item, and the distribution was selected to attempt to equalize for all items the probability of learning during training.

In the random order unlearned (RU) condition, items responded to correctly twice in a row were considered learned by *S*. The next item during instruction was selected from among the items which had not been answered correctly 2 consecutive times.

In the learner-selected (LS) condition, *S* selected the next Japanese word to answer from the appropriate sublist.

The learner-selected known (LSK) condition was the same as LS except *S* was told that previous research had shown it was more effective to practice items he knew to "strengthen their retention" or "try items he thought he knew rather than items of which he was completely unsure," and "to only try items he was unsure of when others were learned very well."

The learner-selected unknown (LSU) condition was the same as LS except *S* was told that previous research had shown that it was more effective to study those items he did not know because "partly known items tended to be almost learned but completely unknown items required quite a bit of practice to get them to the stage where ... was any chance of remembering them."

In all conditions, before instructio... in *S*s typed all 50 Japanese words and En... words from a list provided him.

The retention test, conducted approxi... 1 wk. later, was the same for all *S*s and consist... ittoed sheets with all 50 Japanese words studi... he *S*s were instructed to write the correspond... nglish term for each.

Results. The results of the experime... summarized in Table 1. The first 5 colum... of the table present the proportions correct du... learning; the sixth column presents data from ... retention test. The data from the instructio... ession are presented in 5 successive blocks of 50 tr... each; for the R condition this means that each ...em was presented once in each of these blocks. ...e table gives point estimates of the parameters ...volved. All possible pair-wise comparisons amo... group means were calculated with an overall ...or rate set at .05 within measures by means of th... Bonferroni inequality procedure using the *t* statist... (Miller, 1966). The pooled within-cell variance... or Trial Blocks 1–5 and the retention test were ...5, .010, .029, .041, .047, and .032, respectively. T... following conclusions are based on those compa...ons that indicated statistically significant differen...

In terms of average proportion corre... during training, the conditions may be ordered ... = LSK > LS = RD > RU = LSU. In the LS, ...K, and LSU conditions, the error rates during in...ruction follow the anticipated pattern, indicating ...ost *S*s followed their instructions.

The order of the experimental groups on the retention test is nearly the reverse of that of errors during training. The pair-wise comparisons on the retention scores lead to the order LSU = RD = RU > LS > R > LSK in instructional effectiveness.

The RD, RU, and LSU conditions involve a procedure that attempts to identify and instruct on those items that have not yet been mastered and therefore should produce high error rates during the instructional session.

The magnitudes of the effects observed on the retention test are large enough to be of practical significance. The LSU condition (when compared to the R condition) leads to a point estimate of relative gain of 112.5% in terms of items correct on the retention test. It is clear that *S* can be very effective in determining an efficient study sequence, as effective as the adaptive instructional system used here, and can be made more effective by appropriate instructions.

Discussion. The results show that learning of meaningful paired associates can be facilitated to a considerable degree by some simple instructions to the learner.

On the other hand, it may also be argued that this task does not test the learner-controlled notion very effectively. Learners were probably primarily following the instructions they were given. Factors which have been claimed to affect learner-controlled

sequences such as motivation, learner inquiry, and the use of learner judgment about his own understanding have relatively little opportunity to be effective in the present task.

As in Atkinson's (1972b) study, there is an inverse relationship between proportion correct during learning and proportion correct during testing. This suggests some caution in the popular custom of using trials to criterion or number of errors during instruction as measures of learning in paired-associate studies, particularly when the instructional procedure is sensitive to responses by the learner.

The LSU, RD, and RU strategies were not distinguished on the retention score; the confidence intervals associated with the contrasts were $.05 \pm .18$, $.06 \pm .18$, and $.01 \pm .18$, indicating that whatever differences there may be among these 3 are small compared with the differences between these strategies and the others tested.

Atkinson's (1972b) finding that automated procedures can facilitate learning to a degree of considerable practical significance is supported by the present study. But this study also implies that appropriate instructions to learners can also facilitate learning to a degree of considerable practical significance. Presumably we will be able to design automated instructional devices which are more effective than procedures which the learner himself can implement; but they will have to be more effective at inferring the degree of learning of individual items, and selecting appropriate instructional experiences and sequences, than the learner himself when the learner has the benefit of as much knowledge about the task as that used to design the automated device.

This study does not contradict Atkinson's (1972a) conclusions that learner control may be one source of information in making instructional decisions, but complete dependence upon learner judgment is not warranted. Learner judgment may be an important component in instructional models. Obviously it expends time and other resources, but the exercise of learner judgment may itself result in learning; and this study demonstrates that learner judgment can be used to obtain more efficient instruction than random selection of instructional sequence. Instructional models will no doubt be developed which use learner judgment as a part of the information upon which instructional decisions are based, but it remains to be seen whether the costs associated with this procedure exceed the gains.

This study shows that learners can be taught to more effectively use learner judgment for instructional decisions.

REFERENCES

Atkinson, R. C. Ingredients for a theory of instruction. *American Psychologist*, 1972, **27**, 921–931. (a)

Atkinson, R. C. Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, 1972, **96**, 124–129. (b)

Miller, R. G. *Simultaneous statistical inference*. New York: McGraw-Hill, 1966.

(Received January 10, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 170–172

CHOICE AS A DISRUPTER OF PERFORMANCE IN PAIRED-ASSOCIATE LEARNING[1]

LAWRENCE C. PERLMUTER[2]

Bowdoin College

RICHARD A. MONTY

Human Engineering Laboratory, Aberdeen Proving Ground, Aberdeen, Maryland

PETER M. CROSS

Bowdoin College

In 2 experiments, half of the *S*s were permitted to choose their responses from alternatives following the 1971 procedure of Perlmuter, Monty, and Kimble. The other *S*s were forced to learn the responses chosen by their yoked masters. Both groups of *S*s were forced to learn an interposed A–C list following the choice–force procedure but prior to learning the A–B list. The performance of force *S*s was reliably superior to that of choice *S*s on A–C; however, when both groups learned A–B, the typical facilitative effect of choice was disturbed. In a second experiment, interposing a C–D list produced equivalent performance in both groups on C–D. Similarly, the facilitative effect of choice was not observed on the subsequent A–B trials. The degraded performance of choice *S*s was discussed in terms of frustration as an additional source of motivation.

Allowing *S*s the opportunity to choose a portion of the response materials to be learned on a subsequent paired-associate list was examined with a paradigm that manipulated both the amount of choice allowed during the selection procedure and the locus or point at which choice took place (Monty, Rosenberger & Perlmuter, 1973). The amount of material chosen is not as important as the point at which choice is exercised, as indicated by the finding that even when only 3 out of 12 items were chosen, performance on the entire list was nearly as good as that when all 12 items were chosen, provided that choice was allowed early in the experiment. Such an observation ruled out the associative interpretation of the choice vs. force phenomenon as suggested by Perlmuter, Monty, and Kimble (1971) and seemed to favor a motivational explanation. Subsequently, Perlmuter and Monty (1973) found some preliminary evidence that motivational factors may even work in a detrimental fashion under certain conditions.

A rather extensive literature exists on the relationship between motivational level and performance. It is generally asserted that performance increases as motivation increases. However, excessively high levels of motivation have also been implicated in a deterioration of performance (Brown, 1961). Thus, the purpose of the present experiments is to study further the role of choice in learning and to examine the possibility that choice may have an inhibitory as well as a facilitatory effect on performance.

Experiment I

Employing the basic design of Perlmuter et al. (1971), half of the *S*s chose their A–B responses (choice group), while the other *S*s were forced to learn those responses chosen by their yoked masters (force group). In contrast to earlier experiments, an A–C list was forced upon both groups prior to the learning of A–B, with the expectation that inhibitory effects would cause the choice group to be inferior to the force group on A–C trials, while motivation produced by prior choosing would cause them to be superior to the force group on the subsequent A–B trials. Thus, the analysis of the A–B trials should serve to determine the durability of the beneficial effect of choice.

Method. The *S*s were 40 male and female University of Delaware students, aged 19–25 yr., who were paid $2.00 each for participating in the experiment. They were alternately assigned to 2 groups of 20 *S*s each in order of their appearance.

The apparatus and general procedures were identical to those employed in previous studies. Briefly, the last 84 5-letter words were taken from a list generated by Taylor and Kimble (1967) and were considered to be of relatively high meaning. These items were assembled into 12 groups of 7 each. Within each group, one item was designated arbitrarily as the stimulus word and one item as the A–C response word. The remaining 5 items served as the 5 potential response words on the A–B list. Each *S* was first presented with 3 paired-associate (PA) practice trials utilizing common words. The choice group was then presented with 12 slides, one at a time. Each slide contained a stimulus word centered on the left and 5 potential response words listed vertically on the right. The *S* was instructed to read the stimulus word and the 5 potential response words aloud and to then select

[1] This research was supported in part by Grant MH-15315 from the National Institute of Mental Health to the first author. Experiment I was presented at the meeting of the Southeastern Psychological Association, New Orleans, May 1973.

[2] Requests for reprints should be sent to Lawrence C. Perlmuter, who is now at the Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061.

one word to serve as the response word to the particular stimulus in a subsequent PA learning task. Each S went through the 12 slides at his own rate.

Immediately upon completing the selection of the response items, Ss were informed that they would be required to learn an A–C list prior to the A–B list. The A–C trials were comprised of the stimuli viewed during the selection process, and these stimuli were paired with the E-selected response words indicated above. The stimulus and stimulus-response slides were shown for 2 sec. each. The interval between slides was approximately .9 sec. The S was instructed to say the appropriate word aloud upon presentation of each stimulus slide, and E recorded S's verbal responses.

There were 10 presentations of the 12-word A–C list in 3 different random orders. Each order was presented on every third trial, and each S viewed all items in the same order. A brief period of about 2 sec. followed every third presentation of the list to enable E to return the slide tray to its starting position.

The A–C list was then followed by 10 PA learning trials, with pairs comprised of the stimulus words along with S-selected responses. This A–B list (also presented in 3 different random orders) followed the A–C list after a delay of approximately 50 sec., which was necessary to change slide trays on the projector. The Ss were informed about the change in the list just prior to the A–B trials.

The force group was treated identically to the choice group except that, although they read the stimulus and potential response words aloud, they were not given the opportunity to choose their own responses. Rather, following their reading, they were assigned the response words chosen by the previously tested choice S while each slide was still visible to them. During the choice–force procedure, both groups viewed the slides for approximately 15 sec.

Results and discussion. The number of words correct per trial served as the measure of performance on all analyses. The data for the first A–C trial were omitted from this analysis of variance as all scores were zero. Both trials, $F\ (8,\ 304) = 149.00$, $p < .001$, and the Trials × Groups interaction, $F\ (8,\ 304) = 2.35$, $p < .05$, reached significance, while the groups effect did not. The data underlying the interaction are shown in Figure 1. It can be seen that the force group performed consistently better than the choice group, and that the magnitude of the difference tended to increase generally as a function of trials. This finding is in sharp contrast with the earlier findings (Monty & Perlmuter, 1972; Monty et al., 1973; Perlmuter et al., 1971), all of which indicated that when A–B learning immediately follows the choice procedure, the choice group performs significantly better than the force group, presumably as a result of enhanced motivation stemming from having had the opportunity to choose. However, these results are supportive of Perlmuter et al. (1971), who found

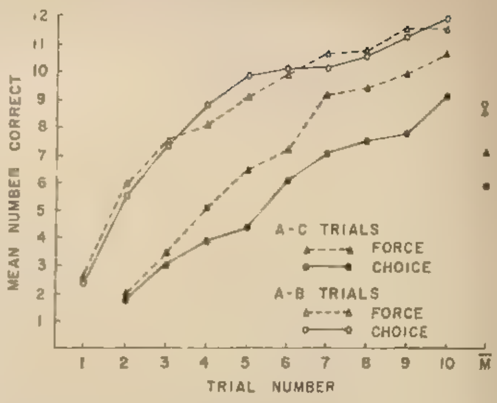


FIGURE 1. Mean number of correct responses on the A–B and A–C trials for the choice and force groups (Experiment I).

that performance on a nonchosen A–C list that was learned following A–B was poorer for choice Ss than for force Ss.

Theoretically, it may be possible to account for the degraded performance of the choice group on the A–C trials in terms of a nonoptimal increase in motivation caused by the interposition of the A–C trials. That is, the additional source of motivation might be thought of as frustration or thwarting (Brown, 1961) caused by the forced learning of the nonchosen A–C list following an opportunity to choose the A–B responses. For force Ss, neither the A–B nor the A–C lists are chosen; hence this frustrative source of motivation should not manifest to the degree that it does for choice Ss. Intrusion data provide further evidence for this interpretation, since increased motivation presumably increases the frequency of competing responses. The choice group made reliably more intrusions (311) than the force group (145), and 87% of the intrusions in each group were misplaced A–C responses. In addition, the finding that choice Ss tend to perform more poorly whether the nonchosen A–C list follows (Perlmuter et al., 1971) or precedes the chosen A–B list lends generality to the proposal that forced learning induces a nonoptimal level of motivation and hence degrades performance.

Finally, when both groups begin to learn the A–B list (following A–C), it would be expected that the incremented motivational level of the choice group would be reduced to a more favorable level. Presumably, the beneficial effect of choice would manifest itself on A–B. In view of the fact that there were differences between groups on the A–C list, however, this analysis must be treated with caution, as discussed by Perlmuter et al. (1971). Only the main effect for trials reached significance, $F\ (9,\ 342) = 142.24$, $p < .001$, indicating a general improvement in performance across trials. As is evident from an examination of the upper 2 curves in Figure 1, the performance of the choice and force groups was highly similar across trials. Thus, it appears as if the beneficial effect of choice noted in previous investigations is destroyed when an

intervening forced list is interposed prior to the learning of A–B. Whether the A–C list enhances the performance of the force group or diminishes the performance of the choice group cannot be ascertained from this experiment.

Experiment II

We have just seen evidence which suggests that A–C trials interpolated between choosing A–B responses and learning the A–B list lead to a deterioration in performance. The nature of the interpolated materials to be learned may be important in the production of this effect. That is, it is entirely possible that an interpolated C–D list may not produce the same degree of frustration that an A–C list might produce. This may be attributable to the degree of formal similarity between A–B and A–C as contrasted with A–B and C–D. Whether the relationships just discussed also determine the performance of the choice and force groups on the subsequent A–B trials will have to await empirical test. Thus, the purpose of Experiment II was to repeat Experiment I, substituting a 10-item C–D list for the A–C list employed in the latter.

Method. The *S*s were 40 male and female students from Bowdoin College. The apparatus and procedures were generally identical to those of the first experiment except that the A–B list was a 12-item high-meaning list similar to that employed in Experiment I, while the C–D list was a 10-item low-meaning list. The A–B list was generated from 72 of the last 75 stimuli on the Taylor and Kimble (1967) list, while the low-meaning C–D list was selected from the first 102 stimuli of the same list. As opposed to Experiment I, 12 rather than 10 trials of each list were presented.

Two groups of 20 *S*s each were run. A C–D choice group was treated identically to the choice group of Experiment I except that following the A–B choice procedure, choice *S*s were required to learn the C–D list, after which they learned the original A–B list. The *S*s in this group were informed of the fact that they would learn A–B (as in Experiment I) after learning C–D. The C–D force group also learned the C–D list before learning the A–B list and was forced to learn the A–B response chosen by the previous *S* in the C–D choice group.

Results and discussion. Performance on the interposed C–D trials (2–12) by the C–D choice group and C–D force group was evaluated. Trial 1 was eliminated from the analysis because there were no correct responses. Only the trials effect reached significance, F (10, 380) = 69.96, $p < .001$, while groups (mean number of correct responses per trial = 2.50 for the C–D choice group vs. 3.00 for the C–D force group) and the Groups × Trials interaction were clearly not significant.

To test the effect of the interposition of the C–D list on A–B learning, performance of the C–D choice group vs. the C–D force group was compared across the 12 A–B trials. The C–D choice group averaged 10.06 correct responses per trial, while the C–D force group averaged 10.08 correct responses per trial. Again, only trials reached significance, F (11, 418) = 87.79, $p < .001$, and the F values for groups and the Groups × Trials interaction were both less than 1.00.

Thus, the results of the present experiment are important in clarifying 2 major points raised by Experiment I. First, when a nonchosen list is imposed upon choice *S*s following their selection of A–B responses, it appears that the similarity of the chosen and nonchosen materials is a critical factor in producing the frustrative effects presumed to have been engendered by the forced list. This conclusion derives from the fact that an interpolated C–D list is learned equally well by C–D choice and C–D force *S*s immediately following the choice-force procedure, which is in marked contrast to the superiority of force *S*s (Experiment I) on A–C. Second, the interposition of a C–D list is sufficient in disturbing the superiority of choice *S*s on A–B. This result is important in substantiating the similar effect observed in Experiment I, in which an A–C list was interposed, but does so under conditions in which first-list learning was not different in the groups subjected to the choice or force procedure. Of course, it is possible that the passage of time is responsible for the equalization of performance on the A–B trials rather than the effect resulting from the interposition of either the A–C or C–D list. This alternative cannot be ruled out completely; rather, it suggests the necessity of determining the temporal stability of the often observed beneficial effect of choice.

References

Brown, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

Monty, R. A., & Perlmuter, L. C. The role of choice in learning as a function of meaning and between- and within-subjects designs. *Journal of Experimental Psychology*, 1972, **94**, 235–238.

Monty, R. A., Rosenberger, M. A., & Perlmuter, L. C. Amount and locus of choice as sources of motivation in paired-associate learning. *Journal of Experimental Psychology*, 1973, **97**, 16–21.

Perlmuter, L. C., & Monty, R. A. Effect of choice of stimulus on paired-associate learning. *Journal of Experimental Psychology*, 1973, **99**, 120–123.

Perlmuter, L. C., Monty, R. A., & Kimble, G. A. Effect of choice on paired-associate learning. *Journal of Experimental Psychology*, 1971, **91**, 47–53.

Taylor, J. D., & Kimble, G. A. The association value of 320 selected words and paralogs. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 744–752.

(Received January 17, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 173–175

MNEMONIC TRANSFORMATIONS AND VERBAL CODING PROCESSES[1]

RAYMOND W. KULHAVY[2] AND JAMES R. K. HEINEN

Arizona State University

Verbal coding research has generally concentrated on parameters of specific mnemonic processes. It appears that many such devices can be explained in terms of some smaller set of control transformations required to accomplish them. Learning of a categorized noun list under serial and total presentations was compared across 4 types of mnemonic coding instructions: imagery, clustering, narrative, and learn. The imagery group had the highest total recall over 3 trials, in spite of clustering *S*s' superior organization. Serial presentation generally enhanced learning and was highly facilitative for the imagery serial group. Results supported the contention that effects of mnemonic instructions on learning are a function of the type of coding transformation necessary to produce them.

Studies of verbal coding processes have concentrated on identifying the parameters of specific control procedures such as imaging, language mediators, prose chaining, and organizational strategies. Essentially, it appears that the majority of known coding mechanisms can be expressed in terms of some smaller set of functional transformations necessary to produce them. Furthermore, one can argue that not only are these transformations necessary for encoding to occur, but they are, in fact, the primary vehicle for transfer of verbal input into long-term memory.

There seem to be at least 2 types of transformational controls that operate when encoding takes place. The first of these we have labeled a *unit* transformation. Unit transformations occur when *S* defines the functional characteristics of verbal stimuli by direct elaboration. For example, if the stimulus is a specific word, *S* might elaborate by placing the word in a visual representation (an image), adding additional language components directly to the word (syntactic mediation), or embedding the word in some semantic context (a narrative).

We will call the second process an *order* transformation. Order transformations accomplish encoding by performing sequence and grouping changes on the input stimuli. In other words, functional characteristics become class dependent. Studies of category clustering are exemplars of this transformation mode.

Of the 2 transformation processes hypothesized, order transformations seem to be the more complex. With unit transformations, *S* simply selects the most available and salient coding vehicle, embeds the new materials within it, and stores the resulting fused compound in memory. In the case of order transformations, however, 2 types of coding information are required for efficient storage. First, *S* must select some primary storage classification (category, superordinate, etc.), and second, he must rearrange the input to match the class chosen. This hypothesis is buttressed by research showing higher recall levels for *S*s who receive category classes as a part of the input information (Tulving & Pearlstone, 1966), allowing a bypass of the first step in the process. The more complex order transformations would, of course, entail treating entire sets of transformed elements as single components in some new ordering strategy.

On the basis of our reasoning, we would predict that instructions to use less complex unit transformations should lead to higher recall scores than instructions to use more complex unit or order transformations—even when the words learned consist of well-defined category sets. If the complexity hypothesis is valid, measures of order transforming (i.e., clustering) should be significantly greater for order-instruction *S*s, in spite of inferior total recall performance.

Finally, a test of the transformation notion would result if stimulus presentation mode were varied. When items are presented singly, unit transformations should be facilitated simply because each item can be processed and stored individually. However, when the list is given en masse, one would predict an advantage for order-transformation instructions, since ease of selecting appropriate primary coding classes would be increased.

Method. Two factors, preexperimental instructions (imagery, clustering, narrative, and learn only) and presentation mode (serial or total) were combined factorially to form 8 treatment groups. Trials were varied as a within-*S*s variable across each factorial cell. The design was thus a 4 (Instructions) × 2 (Presentation Mode) × 3 (Trials) mixed analysis of variance, with repeated measures on the Trials variable.

The *S*s were 83 undergraduate volunteers, randomly assigned to the experimental conditions in order of their appearance for the experiment. Three *S*s were dropped from the study for failure to follow procedural instructions.

The stimuli were 20 nouns selected from the Battig and Montague (1969) category norms. Five

[1] This research was performed under Arizona State University Faculty Research Grant 7805-837-15-8 to the first author.

[2] Requests for reprints should be sent to Raymond W. Kulhavy, Arizona State University, 325 Payne Hall, Tempe, Arizona 85281.

TABLE 1

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) FOR ITEM RECALL

Trial	Instructions							
	Imagery		Clustering		Narrative		Lea[illegible]	
	Serial presentation	Total presentation	Serial presentation	Total presentation	Serial presentation	Total presentation	Serial presentation	[illegible]tion
1	9.41	10.00	8.20	9.11	8.78	[illegible]	[illegible]	[illegible]
	(2.79)	(1.33)	(2.09)	(2.18)	(3.62)	(2.36)	(3.20)	[illegible]
2	15.61	15.23	13.73	13.35	14.53	12.62	12.01	[illegible]
	(1.77)	(2.04)	(3.05)	(2.62)	(3.58)	(2.22)	(3.65)	[illegible]
3	18.10	17.82	15.96	16.31	16.21	15.47	15.24	[illegible]
	(1.10)	(1.87)	(2.02)	(2.16)	(2.74)	(1.83)	(2.97)	[illegible]

Note. Each mean is based on the recall performance of 10 *S*s.

words were chosen from each of the following 4 categories: insects, natural earth formations, seasonings, and sciences. All words were selected from between the third and fourth associative-frequency quartiles for each category. The 20 words were assigned randomly to list position, and were typed on either 3 × 5 in. filing cards or on white foolscap as a triple-spaced list.

During the experimental sessions, *S* sat across a table from *E* and was read verbal instructions appropriate to his treatment. Following instruction, *S* was shown 3 sample items, either serially or together, as an example of the task required.

Imagery *S*s were told to form mental images of each word; clustering *S*s were asked to group the 20 items into common categories; narrative *S*s were directed to chain the words together sequentially in a "short story" that was meaningful for them; and learn-only *S*s merely said each word aloud during presentation.

For serial *S*s, *E* exposed each of the 20 cards at a 3-sec. presentation rate. Speed of exposure was controlled by beeps from a tape recorder. In the total list conditions, *S* was presented with the entire list for 60 sec. Again, time was counted from the taped beeps. To eliminate direct recall from short-term memory, *S*s were required to solve a page of complex arithmetic problems following presentation. The interpolated task was continued for exactly 1 min. Finally, *S* was given a sheet of lined paper and a 2-min. period to free recall as many of the 20 nouns as possible. This procedure was replicated exactly for each of the 3 trials. Following completion of Trial 3, *S* was given a questionnaire assessing the degree to which he had used the appropriate learning strategy.

Recall protocols were scored for total recall, with minor spelling errors counted as correct. Degree of organization was calculated for each recall, using Bousfield's (1953) ratio of repetition (RR). Perseveration errors were discounted in the calculation of RR.

Results. Table 1 presents item-recall means and standard deviations for instructional conditions and presentation modes for each of 3 trials. A 4 (Instructions) × 2 (Presentation Mode) × 3 (Trials) analysis of variance yielded significance for the main effects of instructions, F (3, 72) = 4.63, $p < .01$, and Trials, F (2, 144) = 580.88, $p < .01$, and the Instructions × Trials interaction, F (6, 144) = 2.53, $p < .05$. Comparisons among the instruction means using Scheffé's test showed that imagery *S*s recalled significantly more words than the other treatment groups ($p < .05$). Recall performance of learn-only *S*s was significantly below that of the 3 transformation conditions ($p < .05$). The Instructions × Trials interaction occurred simply because the 3 transformation groups showed a negative acceleration in recall gain for Trials 2 and 3, whereas performance of the learn-only controls was linear due to a lack of ceiling effects. Presentation mode failed to contribute significantly to recall performance.

To assess the degree of order transformation, RR was calculated for each *S* on every trial. Table 2 gives RR means and standard deviations for the instruction and presentation mode conditions across trials. A 4 (Instructions) × 2 (Presentation) × 3 (Trials) analysis of variance was significant for the main effects of instructions, F (3, 72) = 11.57, $p < .01$, presentation mode, F (1, 72) = 5.28, $p < .05$, and Trials, F (2, 144) = 17.00, $p < .01$, and the Instructions × Trials F (6, 144) = 29.00, $p < .01$, and the Instructions × Presentation Mode × Trials, F (1, 144) = 5.50, $p < .01$, interactions. A Scheffé test among the instruction means showed clustering *S*s having a significantly higher RR than the remaining 3 groups ($p < .05$). No other post hoc comparison reached significance. The Instructions × Trials interaction shows that neither the learn-only nor narrative group increased RR appreciably over trials. Imagery *S*s demonstrated a slow rate of climb, and clustering *S*s show a negative acceleration in Trial 3, probably due to both ceiling effects and the restrictive parameters of RR itself (Shuell, 1969).

Discussion. These data are essentially supportive of our model. The learning of a discrete noun list is

TABLE 2

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) FOR RATIO OF REPETITION

Trial	Imagery		Clustering		Narrative		Learn only	
	Serial presentation	Total presentation	Serial presentation	Total presentation	Serial presentation	Total presentation	Serial presentation	Total presentation
1	.21	.14	.37	.41	.23	.20	.26	.10
	(.02)	(.02)	(.05)	(.05)	(.02)	(.03)	(.02)	(.02)
2	.41	.22	.57	.48	.32	.23	.25	.16
	(.03)	(.02)	(.04)	(.05)	(.03)	(.04)	(.02)	(.04)
3	.52	.31	.57	.54	.31	.20	.30	.28
	(.04)	(.04)	(.05)	(.05)	(.02)	(.04)	(.02)	(.04)

facilitated by instructions to transform the items in a specific manner prior to storage. Furthermore, the recall power of such instructions can be predicted from the type of transformation required. The imaginal unit manipulation led to greater facilitation than did either the more complex narrative or order transformations. This finding occurred in spite of the superior organization scores of clustering *S*s.

Differences in item recall as a function of presentation mode did not occur as predicted. It appears possible that input variables are not as important for transformation processes as we assumed. It may be that *S*s tend to treat total list items as single units—transforming them sequentially rather than selectively.

The Instructions × Presentation Mode × Trials interaction provides some interesting results. The reversal of the clustering group performance across trials is clearly due to serial *S*s' lack of an efficient order strategy on the first trial. However, on later trials, organization surpassed that of even the total list group once categories were learned. A more critical point concerns the organization performance of imagery *S*s who received the serial list. The high RR scores for these *S*s on Trials 2 and 3 suggests a tendency to use *both* unit and order transformations to accomplish encoding. Data from the post-experimental questionnaire indicate that these *S*s first transformed the words in a unit fashion, using images, and then began to link the units together, using an order transformation. For example, one *S* in the group responded "first tried to form pictures . . . then saw them together (like) a *katydid* fighting a *mite*. . . ." Here, the unit-transformed images are interactively combined under the category "insects." Several *S*s reported attempts to link the words in some fashion once they had formed the images. Such multiple transformations should result in a high probability of images in the same category being recalled together—exactly the effect shown by our data. In this case, serial presentation may have expedited the multiple-transformation process in an effort to code rapidly or to reduce the load on memory.

The results of this preliminary study are highly suggestive that research on the transformation process will prove profitable. If we are able to isolate the effects of transformations as they occur singly and together in natural language, we will have gone far in explaining the processes underlying human verbal behavior.

REFERENCES

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt. 2).

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229–240.

SHUELL, T. J. Clustering and organization in free-recall. *Psychological Bulletin*, 1969, **72**, 353–374.

TULVING, E., & PEARLSTONE, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

(Received January 22, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 176–177

TIME-OF-OCCURRENCE CUES FOR "UNATTENDED" AUDITORY MATERIAL

STUART T. KLAPP[1] AND PATRICIA LEE[2]

California State University, Hayward

When *S*s shadow (repeat aloud) an auditory message arriving in one ear, memory for material arriving in the other ear falls off over a few seconds of presentation-to-recall delay. Contrary to Glucksberg and Cowen, however, *S*s can accurately estimate the length of delay from the time of presentation of an item in the nonshadowed ear to the time at which the recall signal occurs.

In the shadowing experiment, *S* repeats aloud (shadows) what is heard in one ear while additional material is presented in the other ear. Much interest has surrounded the fate of the material that is not shadowed—the so-called "unattended" material. Although there seems to be little if any long-term memory (LTM) for this material, as assessed by recall at the end of a period of shadowing, short-term memory (STM) may be present (Glucksberg & Cowen, 1970; Norman, 1969). The present investigation focused on one alleged property of STM for nonshadowed material—the possible lack of time-of-occurrence information. Glucksberg and Cowen had *S*s recall digits that were embedded in prose and presented to the nonshadowed ear. A signal light indicated that *S* was to stop shadowing and attempt to recall the digit. The presence of STM without appreciable LTM was shown by a rapid decrease in digit recall as the digit-to-signal delay increased over a few seconds. The *S*s were also asked to "estimate the distribution of delays between digit occurrences and onsets of the cue light [p. 154]" and they "uniformly reported that there had been no delays; either a digit had occurred contiguously with or just before a cue light, or the cue light had flashed alone [p. 154]." This finding led the authors to conclude that memory for nonshadowed material does not contain "time tags." This is indeed a remarkable property for STM and would seem to justify further investigation. However, one must view this conclusion with caution, since *S*s were asked to estimate these delays in terms of a distribution for the entire session, rather than individually for each trial. The purpose of the present experiment was to assess in a more sensitive manner whether *S*s can judge the length of the digit-to-cue delay.

Method. The present experiment attempted to replicate the Glucksberg and Cowen (1970) procedure in all essential features, except that *S*s were asked to estimate the length of the delay from digit presentation to recall cue after each trial rather than to estimate the distribution of these delays at the end of the experiment. It was assumed that this change in procedure would yield a more sensitive estimate of the extent to which *S*s were aware of the time at which the memory digit was presented.

Six right-handed students who reported normal hearing and English as their native language served as *S*s in order to fulfill an introductory psychology course requirement at California State University, Hayward. Before serving in the main experiment, each *S* received 1 min. of practice in shadowing prose in the presence of a distracting message in the other ear.

The message to be shadowed was a passage of English fairy tales, while the nonshadowed message was taken from *David Copperfield* and contained embedded test digits. Both 32-min. messages were read in the same female monotone at 160 words/min. For half of the *S*s, the material to be shadowed appeared in the right ear. The timing relationship between the embedded test digits and the cue light was obtained by recording inaudible signal pulses in one channel of a 2-track tape recorder such that these pulses followed the test digits in the other track by predetermined intervals. These signal pulses triggered a white cue light which was *S*'s signal to stop shadowing and recall the digit. Test digits were randomly selected from the digits 1–9, eliminating the 2-syllable digit 7. All other digits and digit homonyms were eliminated from both prose passages.

Trials were presented in 2-min. blocks, with each block of trials containing all 6 delays (0, 1, 2, 4, 8, and 16 sec.) presented in a random order. The delay from the recall cue of the preceding trial to the introduction of the next test digit was approximately 15 sec. Each *S* received 16 blocks of 6 trials. The *S*s were instructed to stop shadowing when the cue light appeared and to recall any digit recently heard. Although *S*s were encouraged to guess what the digit was, they were not required to respond on each trial. They were also to estimate the delay time from digit to signal light by calling out an estimate of 0–16 sec. each time a digit was recalled.

Results. In agreement with Glucksberg and Cowen (1970), digit recall was approximately 25% at minimum cue delay and decreased at longer delays, $F\ (5, 25) = 2.61$, $p < .05$. As is clear from Table 1, the asymptotic level of recall was higher in the present experiment, but this may be attributable to differences in chance levels in the 2 experiments (S. Glucksberg, personal communication, October 1972). In the present experiment,

[1] Requests for reprints should be sent to Stuart T. Klapp, Department of Psychology, California State University, Hayward, California 94542.
[2] Now at Michigan State University.

smaller set of digits was used, and there were no "blank" trials in which a cue light appeared when there had been no digit. Unfortunately, it is difficult to assess the chance level in either experiment, since *S*s were not required to respond on each trial. Such a procedure would have been useful for purposes of analysis, but would seem unnatural to *S*s, who frequently reported that no digit at all had been heard. Regardless of these details, it seems reasonable to conclude that the decaying STM for unattended digits has been replicated in the present experiment, so that it appears that the same basic phenomenon is being investigated.

Considering only those trials on which digit recall was correct, *S*s' estimates of the cue delays were highly correlated with the actual delay from digit to cue light, $r = .90$, $p < .001$. Furthermore, 76% of the estimates were not more than 1 sec. different from the actual delay, while 32% of the estimates were identical to the actual delay.

Discussion. The present results provide no support for the assertion of Glucksberg and Cowen (1970) that time cues are missing from STM for nonshadowed material, but rather demonstrate that time-of-occurrence estimates can be made with great accuracy. Having eliminated the alleged [illegible] of time cues as a property of STM for non-shadowed material, there appears to be no reason [illegible] assume that STM for nonshadowed material is different from STM in the distractor paradigm of Peterson and Peterson (1959). In both cases, recall falls off dramatically after only a few seconds [illegible] activity. The lower initial recall of nonshadowed material compared to the initial recall in the distractor paradigm can be attributed to reduced perceptibility of the to-be-remembered items in the presence of the sound and the verbal [illegible] of shadowing. However, forgetting of perceived items may be assumed to follow the same laws in both cases in the absence of evidence to [illegible].

[illegible]shadowed material seems to be understood even though it is not remembered for long. Corteen and Wood (1972) observed galvanic skin responses to names of cities in the nonshadowed channel when other city names had previously been associated with shock. Lewis (1970) found that the reaction time to repeat (shadow) a word was a function of its semantic relationship to a simultaneously presented word in the nonshadowed message. In both of these studies, there was essentially no recall of nonshadowed material at the end of the trials, indicating that the material had not entered LTM even though it was comprehended at some level. These findings, as well as the rapidly decreasing memory for nonshadowed material, may be handled by assuming that shadowing reduces rehearsal in a manner analogous to the role of counting in the Peterson and Peterson (1959) paradigm. It is commonly assumed that reducing rehearsal reduces LTM. On the other hand, rehearsal may not be needed for comprehension or STM—both of which appear to exist for nonshadowed material.

TABLE 1

Percent Digit Recall as a Function of Cue Delay (in sec.)

Source of data	Delay of cue					
	0	1	2	4	8	16
Glucksberg and Cowen (1970)—estimated from curve	27	20	14	8	5	5
Present experiment	26	23	17	17	14	11

REFERENCES

Corteen, R. S., & Wood, B. Autonomic responses to shock-associated words in an unattended channel. *Journal of Experimental Psychology*, 1972, **94**, 308–313.

Glucksberg, S., & Cowen, G. N. Memory for non-attended auditory material. *Cognitive Psychology*, 1970, **1**, 149–156.

Lewis, J. L. Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 1970, **85**, 225–228.

Norman, D. A. Memory while shadowing. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 85–93.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

(Received January 26, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 178–180

CONTEXTUAL CONSTRAINTS AND THE PERCEPTION OF SPEECH

STEVEN ROSENBERG[2]
Carnegie-Mellon University

WALLACE E. LAMBERT
McGill University, Montreal, Quebec

In the shadowing of connected discourse, *Ss'* performance deteriorated when intersentence relations were altered or disrupted. The following 3 methods were used to disrupt intersentence relations: (*a*) insertion of extraneous material, (*b*) permutation of sentence order, and (*c*) deletion of sentences. The insertion of extraneous material had the greatest effect on shadowing performance. This performance was most disrupted in the passages with the greatest number of anaphoric relations. At the shadowing lags found in the present study, *Ss* used intersentence relations as cues in their speech processing.

The present experiment investigated the effect of varying the contextual constraints between sentences on the shadowing of connected discourse. A shadowing task is of interest primarily because of the light it sheds on the mechanisms of immediate speech processing. Earlier researchers (Chistovitch, 1960; Chistovitch, Aliakrinskii, & Abul'ian, 1960; Galunov & Chistovitch, 1966) found that it was possible to shadow as closely as 200 msec. Consequently it was possible, when pressed, to shadow on the phonemic level, or, at most, the level of the word. In this case, the shadower does not use any other information beyond that in the immediate word or phoneme being processed. More recent experiments indicate that syntactic and semantic information from the whole sentence can be operative in shadowing. Using a series of statistical approximations, Treisman (1965b) found that *Ss* shadowed with less error as the material approached normal English. Using mixed sets, Miller and Isard (1963) found that normal sentences were better shadowed than nonsense sentences, which in turn were more easily shadowed than scrambled nonsense sentences. In our turn, we have shown how the relations between sentences can affect shadowing ability.

Olney and Londe (1966) and Olney (1969) investigated the class of intersentence relations called anaphoric relations, and specified criteria for tracing these relations between sentences. An anaphoric expression is one which functions in part as a substitute for a preceding expression in the same discourse. For example, in the following sentence, "It" is an anaphoric reference: "John hit the ball. It went over the fence." The number of anaphoric expressions in each test passage is a measure of the intersentence relations in that passage, although there are of course many other types as well. Two expressions were considered to be anaphorically related when they were coreferential. This class of intersentence relations is not exhaustive, but it does provide one quantitative measure of the contextual coherence of the 3 test passages, and is indicative of the relative amounts of intersentence relations in them.

Method. Three test passages were chosen which contained different amounts of intersentence connections. Passage I was a passage from an introductory textbook in psychology, Passage II was a newspaper editorial, and Passage III was a descriptive passage from a novel. Passage I had 40 anaphoric relations, Passage II had 21, and Passage III had 17 such relations.

To investigate the effects of the disruption of *contextual coherence* on shadowing ability, 8 levels of *contextual connectivity* (including the original unaltered passages) were established by systematically altering the 3 test passages. Four of these 8 levels are defined as follows: Level 1, original versions; Level 2, random-order sentences, in which each passage had the order of its sentences randomized, disrupting order- and contiguity-dependent intersentence relations; Level 3, reversed order of sentences, in which closeness of each sentence to the other was maintained so that, for example, word associations might still hold, but the sequence of order-dependent relations was altered; and Level 4, random sentences from the same source, in which new passages were formed by choosing, in each case, 22 sentences (without duplicating earlier choices) at random from the same source that provided the original passage; this preserves style, but little else.

The following manipulations, unlike the preceding ones, preserve the sentence order, but introduce extraneous material: In Level 5, alternation of passage sentences with samples from same source, a passage was chosen from the same source as the original and its sentences were alternated with the sentences from the original passage, in order, so that the first sentence from the first passage was followed by the first sentence from the second passage; then the second sentence from the first passage, and so on This intersperses the test passage with a second coherent passage in the same style. In Level 6, alternation of passage sentences with samples from other sources, sentences from Passage I were alternated with sentences from Passage III, likewise the sentences from Passages II and III, and from Passages I and II. This is similar to Level 5, except that now the interspersed

[1] This research was supported by subventions from the Canadian Defense Research Board and the Canada Council.
[2] Requests for reprints should be sent to Steven Rosenberg, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

passages differ on such things as style and a more divergent vocabulary. Level 7, alternation of partial passages from the same source, was identical to Level 5, except that every second sentence in each passage was deleted. Hence the first sentence of Passage I was followed by the second sentence of Passage II, and then by the third sentence of Passage I, etc. This reduces the coherence of the passage, which might have a greater effect on a passage such as I, which developed an argument, than on Passage III, which was descriptive. Level 8, alternation of partial passages from different sources, was identical to Level 6, except that, as in Level 7, every alternate sentence was deleted. Once again there are now style differences between the partially destroyed passages.

Passages I, II, and III each contained 22 sentences, although the number of words varied. Note that in each of the 8 conditions the test passage sentences have not been altered, only their relation to each other has been rearranged. Two practice passages about 6 min. in length were administered. These were literary in nature, but bore no relation through content or style to the test material.

The resulting 24 passages were recorded by a linguist at an average rate of 150 words/min, using normal inflection. Each *S* shadowed only one test passage, lasting 3–6 min. Five *S*s were tested on each passage, with the exception of the 3 original passages, where 11 *S*s per passage were used. A total of 138 *S*s participated. These *S*s were volunteers who had taken at least one psychology course, making them familiar with the "technical" terms in Passage I. They were tested in groups of approximately 12, in separate booths (with an empty booth between) of a language laboratory. All equipment was controlled by *E*, who also read the instructions aloud. The stimulus passage was given over headphones and also recorded on one track of a tape. The shadowed response was recorded on the other track.

Scoring. The raw data were scored for the number of errors per sentence. These results were transformed into percentages of words correctly shadowed per passage, permitting comparisons across passages. An error consisted of any response which differed, through omission, distortion, or substitution, from the original.

To score the time lag in shadowing, each of 2 scorers listened to a separate track containing either the test passage or the shadowed response. When the scorer listening to the stimulus track heard each fifteenth word, he pressed a relay to start an automatic timer. The second scorer also pressed a relay when hearing the same word on the track containing *S*'s response, stopping the timer. The timer gave an automatic printout of each time interval to the nearest millisecond, which was rounded off to the nearest tenth of a second. There were 30–60 time intervals measured per test passage.

Both error and time scores were taken only on those sentences from the original test passages.

TABLE 1

MEAN PERCENTAGE CORRECT (IN PARENTHESES) AND MEAN TIME SCORES (IN SEC.)

Contextual connectivity level	Passage I (Text)	Passage II (News)	Passage III (Novel)	*M*
1. Original version	(73.5) .681	(75.1) .498	(82.5) .517	(77.1) .565
2. Random order, within passage	(63.8) .792	(68.8) .515	(66.7) .668	(66.2) .659
3. Reversed order, within passage	(78.4) .690	(82.5) .509	(68.9) .529	(76.6) .576
4. Random, same source	(68.3) .564	(56.7) 1.114	(72.3) .611	(65.9) .763
5. Alternated, same source	(47.6) .832	(61.7) .643	(72.8) .544	(60.7) .673
6. Alternated, different source	(59.3) 1.074	(73.5) .742	(61.7) .758	(64.7) .858
7. Alternated and omitted, same source	(74.0) .762	(72.6) .576	(65.2) .651	(70.6) .663
8. Alternated and omitted, different source	(71.4) .769	(70.0) .588	(68.5) .673	(69.9) .677
M	(66.2) .770	(72.1) .648	(68.7) .619	

In conditions where other material was used to intersperse the sentences of the original passage, this interspersed material was not scored. Consequently some passages had a larger number of sentences scored than other passages. For example, in the last 2 conditions, sentences were deleted from the original passages. As there might have been a "ceiling" effect in the error scores, which were percentages, they were reanalyzed using an arc sine transformation (Winer, 1962, p. 221). These analyses did not differ significantly from those performed on the untransformed data. It is the analyses of the untransformed data that are reported below. The time latencies, on the other hand, had a lower, but no upper, limit, hence geometric means of each *S*'s scores were used in computing a mean time lag per *S* for a test passage. Nonhomogeneity of variance tests were not significant.

Results and discussion. Table 1 gives the mean percentage correct and mean time lag for the 3 original passages, and all manipulations of these passages. The 3 unaltered, original passages (Level 1) were tested for differences using a one-way analysis of variance. The analysis did not reveal significant differences in the ability of *S*s to shadow these passages.

Two separate analyses of variance (two-way) were performed on all 8 levels of the 3 test passages, on the error and time lag scores, respectively. There were significant passage differences for time lag scores, $F\ (2,\ 114) = 3.32$, $p < .05$. To investigate these passage differences, a one-way analysis of variance was performed on each test passage over all 8 levels of contextual connectivity. Passage I had significant variations in both error, $F\ (7,\ 38) = 3.96$, $p < .01$, and time scores, $F\ (7,\ 38) = 3.07$, $p < .05$, over all levels. Passage II had significant variations in time scores, $F\ (7,\ 38) = 3.79$, $p < .01$, over all manipulations. Passage III did not produce any significant differences.

To determine which manipulations were causing this effect, each level was compared to Level 1, the original versions. The comparison of Level 1 with Level 5 resulted in significant differences in error scores, t (47) = 2.22, $p < .05$. Similarly, the comparison of Level 1 with Level 6 resulted in significant differences in both error, t (47) = 2.24, $p < .05$, and time scores, t (47) = 3.05, $p < .01$. The results seem concentrated in the manipulations that preserve sentence order, but introduce an alternate context. To test this, a two-way analysis of variance was performed on Levels 1–3, which vary sentence order. No significant effects occurred. Similarly, Level 1 with Levels 5–8 were analyzed, and significant main effects for time and error differences between the levels resulted, Fs (4, 93) = 4.53 and 4.24, ps $< .01$, for the time and error scores, respectively. To test this, Level 1 was contrasted to Levels 5–8. This was significant for the error scores, F (1, 93) = 4.86, $p < .05$.

Our results show that in the shadowing of connected discourse, *S*s may use the relations between sentences. Along with earlier findings (Miller & Isard, 1963; Treisman, 1965a, 1965b) the authors have shown that *S*s are able to use contextual cues in shadowing. We have extended these previous findings that *S*s use relations within the sentence by showing that they are able to use the intersentence relations found in connected discourse as well.

Our manipulations of test passages reduced the ability of *S*s to shadow the *same* sentences in the various conditions. The insertion of sentences from an extraneous passage into the [illegible] assage caused most of the disruption in [illegible] mance. A larger effect was found in passag[illegible] n more anaphoric relations. No differences w[illegible] ound in the ability to shadow the unalter[illegible] ssages. Either the anaphoric relations do no[illegible] ct the shadowing of normal passages, or th[illegible] s differed on some other measure which [illegible] ces out any effect due to different numbers [illegible] aphoric relations.

REFERENCES

CHISTOVITCH, L. A. Classification of rapidly [illegible] sped sounds. *Soviet Physics: Acoustics*, 1960, 5, [illegible] 8.

CHISTOVITCH, L. A., ALIAKRINSKII, V. V., & [illegible], V. A. Time delays in speech repetition. *Problem[illegible] Psychology* 1960, 1, 64–70.

GALUNOV, V. I., & CHISTOVITCH, L. A. [illegible] onship of motor theory to the general problem of sp[illegible] ogaush (review). *Soviet Physics: Acoustics*, 1966, 11, [illegible] 365

MILLER, G. A., & ISARD, S. Some perceptual [illegible] nsequences of linguistic rules. *Journal of Verbal Learni[illegible] and Verbal Behavior*, 1963, 2, 217–228.

OLNEY, J. *An investigation of English disco[illegible] structure, with particular attention to anaphoric relation[illegible]* (Document TM 4229) Santa Monica, Calif.: Systen [illegible] velopment Corporation, March 1969.

OLNEY, J., & LONDE, D. *A research plan f[illegible] nvestigating English discourse structure, with particula[illegible] ttention to anaphoric relationships.* (Document TM ([illegible] 56) Santa Monica, Calif.: System Development Corp[illegible] on, March 1966.

TREISMAN, A. M. The effects of redundancy a[illegible] familiarity on translating and repeating back a foreign [illegible] d a native language. *British Journal of Psychology*, 19[illegible], 56, 369–379. (a)

TREISMAN, A. M. Verbal responses and co[illegible] xtual constraints in language. *Journal of Verbal Learni[illegible] and Verbal Behavior*, 1965, 4, 118–128. (b)

WINER, B. J. *Statistical principles in experi[illegible] tal design.* New York: McGraw-Hill, 1962.

(Received February 1, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 181–183

CONDITIONS THAT DETERMINE EFFECTIVENESS OF PICTURE-MEDIATED PAIRED-ASSOCIATE LEARNING[1]

KEITH A. WOLLEN[2] AND DOUGLAS H. LOWRY[3]

Washington State University

Previous research has demonstrated that paired-associate learning is facilitated when the pairs are accompanied by pictures showing the pair members in some sort of spatial or interactive relationship. Two experiments extended the conditions under which such facilitation occurs. In Experiment I, facilitation was obtained even though all pictures were presented before any of the word pairs and no specific instructions were given to use the pictures. In Experiment II, interacting pictures produced facilitation relative to either a no-picture control or to controls having noninteracting pictures. Facilitation was obtained when the pictures were presented either before the word pairs or at the same time, but not when the pictures were presented after the pairs.

It is well known that *S*s can learn paired associates more rapidly when an interacting picture is shown along with each word pair than when no such picture is presented (e.g., Epstein, Rock, & Zuckerman, 1960; Wollen, Weber, & Lowry, 1972). A common interpretation of such facilitation is that the picture is stored with the word pair during input, and at output, the stimulus term presumably elicits an image of the interacting picture, which is then decoded into the response term (Paivio, 1971). In previous research on such matters, *S*s have been given the pictures at the same time as the words, and have been instructed to use the pictures to help them learn the pairs. The questions raised in Experiment I are: (*a*) just how effective are pictures when they are presented at a different time from the word pairs, and (*b*) whether pictures produce interference if they show 2 objects, only one of which is contained in the to-be-recalled pair.

EXPERIMENT I

Method. The learning materials consisted of the following pairs of concrete nouns: List 1—BOTTLE-HAMMER, KEG-TABLE, NAIL-FLOWER, SNAKE-ELEPHANT, WIGWAM-UMBRELLA, CHAIR-HOUSE, HORSE-TREE, WHALE-BIRD; and List 2—BOTTLE-KETTLE, KEG-APPLE, NAIL-LOBSTER, SNAKE-REVOLVER, WIGWAM-ARROW, CHAIR-FLAG, HORSE-FIRE, WHALE-DIAMOND. The 2 lists were counterbalanced such that half of the *S*s learned List 1 and half List 2. The words ranged from 6.23–6.90 on the imagery scale of Paivio, Yuille, and Madigan (1968). For each pair of words, a silhouette picture was made that depicted the 2 nouns in a spatial or interactive relationship (e.g., an arrow stuck in a wigwam, a hammer about to hit a bottle, etc.). Each silhouette and each word was made into a 2 × 2 in. slide. These materials were projected onto a 60-in. viewing screen.

The *S*s were 100 introductory psychology students whose participation partially fulfilled a course requirement. The *S*s were run in small squads of 5; these squads were randomly assigned to 5 experimental groups. As indicated in Table 1, the groups differed with respect to which of 4 different stages were used and which types of pictures were employed. In Stage 1, designated A–P_A, a word (the A term) was shown by itself for 1.5 sec. and was followed by a picture of that A term (P_A) presented by itself for 1.5 sec. This procedure continued until all eight A–P_A combinations had been presented. The *S*s were told to associate each word with the picture that followed it. Stage 2 consisted of the presentation of 8 pictures without any words. In each picture, 2 objects were depicted in a spatial or interactive manner. Depending upon the condition, the drawings pictured the 2 terms of the to-be-recalled pair (P_{A+B}) of Stage 3, the response term of the Stage 3 pair with an irrelevant object (C) (P_{C+B}), or the stimulus term of the Stage 3 pair with an irrelevant object (P_{A+C}). Each of these pictures was shown for 1.5 sec.; *S*s were told to pay close attention so that they could remember the pictures. Stage 3, designated A–B, involved the presentation of the stimulus word (A) of a pair for 1.5 sec., then the response word (B) for 1.5 sec., and finally a 1.5-sec. blank interval which served to separate the pairs. The *S*s were instructed to learn the pairs so that they could provide one member given the other; no mention was made of using the pictures from the previous stages. Finally, Stage 4 (test) consisted of presenting each A word for 1.5 sec. followed by a 3.5-sec. interval for *S* to write the missing B term.

Performance was expected to be the highest in Groups 1 and 2 since the objects pictured in Stage 2 were the same as those named in the A–B stage. In Groups 3 and 4, in contrast, the pictures might produce interference, as only 1 pair member (either A or B) was depicted in Stage 2. Group 5 represents a no-picture control.

Results. The mean numbers correct and numbers of overt errors are shown in Table 1. Since there were significant differences among the mean numbers

[1] This investigation was supported in part by U.S. Public Health Service Research Grant HD-02249 from the National Institute of Child Health and Human Development. Appreciation is expressed to Marvin Brown, who ran Experiment I.

[2] Requests for reprints should be sent to Keith A. Wollen, Department of Psychology, Washington State University, Pullman, Washington 99163.

[3] Now at Michigan Technological University.

TABLE 1

Frequency Data in Experiment I

Group	Frequency	
	Correct responses	Errors
1 A–P_A, P_{A+B}, A–B, test	6.40	.65
2 —, P_{A+B}, A–B, test	6.05	.45
3 A–P_A, P_{C+B}, A–B, test	3.70	1.90
4 A–P_A, P_{A+C}, A–B, test	3.50	1.70
5 —, —, A–B, test	4.25	1.30

Note. Abbreviations: A = stimulus term, P = picture, B = response term, C = irrelevent object.

correct, F (4, 95) = 55.15, $p < .001$, a Dunnett's test was conducted comparing the Group 5 mean with each of the other 4 means. A difference of 1.23 was required at the .05 level. As expected, Groups 1 and 2 were significantly higher than Group 5. However, the trends for Groups 3 and 4 to score below the control group were not significant. The error data, which show the same trends, were not analyzed due to the small means and skewed distributions. In general, it appears that pictures can enhance performance even though they are presented at a different time from the word pairs.

Experiment II

In Experiment I, the superiority of Groups 1 and 2 might have been attributable to the fact that these groups in effect had 2 trials (P_{A+B} and A–B) whereas Group 5 had only 1 (A–B). This possibility was explored in Experiment II by giving the standard control group 2 successive trials (i.e., A–B, A–B) rather than 1. In addition, for every interacting-picture condition, there was another control group, which received pictures with the objects in a side-by-side, noninteracting fashion. The use of noninteracting pictures equated each interacting-picture group with its noninteracting control group in every respect except for the presence or absence of interaction; hence, interaction per se can be examined. Finally, groups were added in which pictures were presented (*a*) along with the word pairs, and (*b*) after all word pairs.

Method. The materials were the same as those used by Wollen et al. (1972) in their interacting–nonbizarre and noninteracting nonbizarre conditions. Basically, there were 9 high-imagery noun–noun pairs, each of which appeard at the bottom of a slide. When pictures were used, they appeared above the word pairs.

The design consisted of the 9 groups outlined in Table 2. The first and second columns indicate the conditions present on Trials 1 and 2, respectively. The symbols are the same as those used in Experiment I with the exception of P_{A-B} which refers to noninteracting pictures. For example, Group 2 received noninteracting pictures without words on Trial 1 and the word pairs al[illegible]rial 2; Group 6 received interacting pictures [illegible]anied by word pairs on both trials, and [illegible]ch group contained 20 introductory psych[illegible]dents (N=180) whose participation partially [illegible]d a course requirement. The *Ss* were [illegible]mall squads of 5 each. A randomized block [illegible]dure was used in which one squad was run [illegible] of the 9 conditions before beginning the next [illegible]

The sequence of events wa[illegible] Trial 1, 30-sec. interval, Study Trial 2, 50-se[illegible]l, Test Trial 1, 70-sec. interval, Test Trial 2 [illegible] pairs were presented in the same order on th[illegible]y trials and in 1 of 4 different orders on the 2 [illegible] The intervals between trials were used to pr[illegible]structions and to pick up and distribute re[illegible] sheets. Each slide, whether words only, pic[illegible] or both, was presented for 1.5 sec. with [illegible] on study trials. On test trial[illegible] were shown each stimulus word (or a[illegible]d naming one of the pictured objects) for 2 sec., f[illegible]wed by a 4-sec. blank period, during which *S* w[illegible] the missing response word (or object). All m[illegible] were presented via back projection to a 20[illegible]

The instructions varied some[illegible] according to condition. On Study 1, Group[illegible] 2 were told to study the pictures and to re[illegible]member as much about them as they could. Then on Study 2, these *Ss* were told to try to rememb[illegible] picture from Study 1 as an aid to Study 2 [illegible] they were also instructed to learn such that [illegible] could provide one member given the other. Groups 3, 4, and 5 were all told on Study 1 to l[illegible] so they could provide one member given th[illegible]ther. Group 5 received similar instructions for Study 2, but Groups 3 and 4 were instructed to study the pictures because the pictures would help them remember the previously presented word pairs. For both study trials, Groups 6 and 7 were told to use the pictures to help them learn the pairs. [illegible]ally, Groups 8 and 9 were instructed to learn the pictures so that they could name one object when given the name of the other. The instructions for test trials were simply to write the missing pair member.

Results and discussion. The mean number correct was determined for each test trial. Since there was

TABLE 2

Mean Correct in Experiment II

Group	Item		
	Study Trial 1	Study Trial 2	M correct
1	P_{A+B}	A–B	5.98
2	P_{A-B}	A–B	4.58
3	A–B	P_{A+B}	5.40
4	A–B	P_{A-B}	5.65
5	A–B	A–B	4.18
6	A–B P_{A+B}	A–B P_{A+B}	6.72
7	A–B P_{A-B}	A–B P_{A-B}	5.32
8	P_{A+B}	P_{A+B}	6.10
9	P_{A-B}	P_{A-B}	4.88

Note. Abbreviations: P = picture, A = stimulus term, B = response term; P_{A-B} indicates a noninteracting picture and P_{A+B} indicates an interacting picture.

Means are for combined data for the 2 test trials.

relatively little difference between the 2 trials, Table 2 presents only the means of the 2 test trial means. A simple randomized analysis of variance on all 9 groups was significant, F (8, 171) = 3.23, $p < .01$. Consequently, each group was compared to the standard control (Group 5) using Dunnett's test. The difference required for significance at the .05 level (1.51) was only reached by those groups having interacting pictures on Study 1 (Groups 1, 6, and 8). As expected, having interacting pictures only on the second study trial (Group 3) failed to produce a significant amount of facilitation. Similarly, none of the conditions with noninteracting pictures produced a significant amount of facilitation.

In addition to the above analyses, 1-tailed t tests were run comparing each group having interacting pictures with its noninteracting-picture control. Group 1 was higher than Group 2, t (38) = 1.86, Group 6 was above Group 7, t (38) = 2.43, and Group 8 was above Group 9, t (38) = 2.37, all $ps < .05$. However, Group 3 was not above Group 4; in fact, Group 3 was slightly inferior. Hence, the same results emerge as when each group was compared with Group 5; namely, interacting pictures facilitate when pictures are presented on Study 1 (Groups 1, 6, and 8) but not when they appear only on Study 2 (as in Group 3).

Together, Experiments I and II indicate that interacting pictures produce a higher level of response than noninteracting pictures, or than no pictures at all. This facilitation occurs when (*a*) the pictures and word pairs are presented simultaneously or (*b*) all of the pictures are presented before any of the words, but not when (*c*) the pictures are given only after the presentation of the word pairs.

REFERENCES

Epstein, W., Rock, I., & Zuckerman, C. B. Meaning and familiarity in associative learning. *Psychological Monographs*, 1960, **74**(4, Whole No. 491).

Paivio, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Pt. 2).

Thomson, D. M., & Tulving, E. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 1970, **86**, 255–262.

Wollen, K. A., Weber, A., & Lowry, D. H. Bizarreness versus interaction of mental images as determinants of learning. *Cognitive Psychology*, 1972, **3**, 518–523.

(Received February 20, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 1, 184–186

ROLE OF DETAILS IN THE LONG-TERM RECOGNITION OF PICTURES AND VERBAL DESCRIPTIONS[1]

THOMAS O. NELSON[2]
University of Washington

JACQUELINE METZLER
Stanford University

DAVID A. REED
University of Washington

University undergraduates studied and then had immediate and delayed recognition tests on either (*a*) photographs, (*b*) verbal descriptions of the main theme of the photographs, (*c*) unembellished line drawings of the main theme of the photographs, or (*d*) main-theme line drawings embellished with extra details from the photographs. Although the 3 pictorial conditions had higher recognition than the verbal condition, they did not differ from each other on either the immediate or delayed recognition tests. These results imply that the recognition advantage for pictures over verbal descriptions is not due to the extra details which pictures contain.

It is well known that immediate recognition is very high for ordinary pictures (Nickerson, 1965; Shepard, 1967). Furthermore, immediate recognition is nearly always higher for pictures than for verbal descriptions (but see Goldstein, Locke, & Fehr, 1972). It has also been established that *long-term* recognition is quite high for pictures (Nickerson, 1968; Shepard, 1967). Although the long-term recognition of pictures has not been directly compared with that of verbal descriptions, the results of studies examining long-term recognition of verbal descriptions only (e.g., Strong, 1913) suggest that long-term recognition might be higher for pictures than for verbal descriptions. Such a direct comparison in long-term recognition of pictures vs. verbal descriptions was one focus of the present study.

Because of the myriad of details contained in ordinary pictures, the verbal descriptions of most pictures are not entirely complete (cf. the aphorism "a picture is worth a thousand words"). One might explain any recognition advantage for pictures over verbal descriptions in terms of the extra details contained in the pictures but not in the verbal descriptions; if *S* encodes additional pictorial details during study, then recognition of any one of those details at the time of test could serve as the basis for distinguishing the target picture from the distractor. These extra details, collectively referred to here as "embellishment," have been considered by previous researchers (Haber, 1970; Nickerson, 1965) as a possible determinant of the high recognition accuracy for pictures. The second (and major) focus of the present study was to directly manipulate the number of details contained in the to-be-recognized pictures to determine if extra pictorial details facilitate recognition.

Method. There were 4 sets of stimuli, each containing 120 items. One set (PH) consisted of black-and-white photographs used in a previous experiment (Nickerson, 1965).[3] Two student judges independently examined the photographs and generated a one-sentence description of each photograph. The few differences that occurred in the 2 judges' descriptions were reconciled by a third student judge. The final 120 verbal descriptions (VD) comprised the second stimulus set. Line drawings of the photographs were constructed on the basis of the PH and VD stimulus sets, so as to include only those aspects of the photograph that were mentioned in the verbal description. The resulting line drawings comprised the third stimulus set, unembellished line drawings (ULD). Finally, extra details from the PH stimuli were added to ULD to form the fourth stimulus set, embellished line drawings (ELD). It should be noted that ELD was not as detailed as PH, since the latter contained shading and fine-grain details that could not be included in a line drawing. An example of an item appearing in each of the 4 conditions (VD, ULD, ELD, and PH) is shown in Figure 1.

The *Ss*, run in groups of 4–5, were 136 undergraduates at the University of Washington, whose participation fulfilled a course requirement. One independent variable (between-*Ss*) concerned the type of to-be-recognized items. The *Ss* were assigned to conditions according to order of appearance, beginning with 34 *Ss* per condition, and complete data were obtained from 33 *Ss* in Condition PH, 34 *Ss* in ELD, 30 *Ss* in ULD, and 31 *Ss* in VD (the loss of data from 8 *Ss* was due to failure to return for the second session). The second independent variable (within-*Ss*) was the retention interval for the recognition test, which was either immediate (7 min. after the items had been studied) or delayed (7 wk. after the items had been studied). Both tests were 2-alternative forced-choice recognition tests. To avoid the problem of having once-

[1] This research was begun while T. O. Nelson was a National Science Foundation Postdoctoral Fellow at Stanford University, 1970–71, and was completed with the support of U.S. Public Health Service Grant MH-21037 to T. O. Nelson. The authors thank David Gorchoff for his help in pilot research, Colin MacLeod and Karen Woodbury for assisting in the statistical analysis, and Geoffrey Loftus for his comments on the manuscript.

[2] Requests for reprints should be sent to Thomas O. Nelson, Department of Psychology, University of Washington, Seattle, Washington 98195.

[3] Raymond S. Nickerson kindly provided these photographs.

tested items in the immediate test and twice-tested items in the delayed test, the following technique was used. Of the 120 items in a given condition, 60 were studied and 60 served as distractors. Thirty of the 60 studied items were tested only on the immediate test, whereas the remaining 30 studied items were tested only on the delayed test; similarly, 30 of the 60 nonstudied items were the immediate-test distractors, whereas the remaining 30 nonstudied items were the delayed-test distractors. Items were counterbalanced across *S*s so that a given item served approximately equally often in each of the 4 studied–nonstudied immediate–delayed conditions.

During study, a Kodak Carousel projector presented the items at a rate of 10 sec. per item. Following presentation of the last study item, *S* solved arithmetic problems for 7 min. while *E* rearranged the slide trays for the immediate recognition test. During the recognition test, pairs of test items were presented simultaneously from 2 Kodak Carousel projectors, synchronized by an external timing device to present each pair of test items for 15 sec. The *S* made a written response to indicate whether the previously studied item was the left or right member of the test pair, guessing whenever necessary. Half of the previously studied items were on the left and half were on the right, intermixed according to a Gellerman sequence (1933). Seven weeks later, *S* returned for the delayed recognition test. The 30 test pairs were presented at a rate of 20 sec. per pair, and *S* made a written response to indicate which member of the pair he had previously studied, guessing whenever necessary.

Results. The mean percentage of correct recognitions for each condition of the immediate and delayed tests is shown in Figure 2. These data were analyzed by a 4 × 2 (Conditions × Retention Interval) mixed-model analysis of variance. Recognition accuracy was significantly higher on the immediate test than on the delayed test, F (1, 124) = 284.5, $p < .001$. The main effect of the 4 conditions was also significant, F (3, 124) = 13.3, $p < .001$, but only because PH, ELD, and ULD were significantly higher than VD (PH, ELD, and ULD were not significantly different from each other, $p > .05$). The Conditions × Retention Interval interaction was also significant, F (3, 124) = 10.5, $p < .001$; analyses of the simple interaction effects showed that recognition accuracy declined over the retention interval significantly more slowly for PH, ELD, and ULD than for VD (the rates of decline for the former 3 conditions did not differ significantly, $p > .05$). Parenthetically, it should be noted that, except for a significant simple interaction effect of ELD–ULD × Retention Interval ($.025 < p < .05$), transformation of the percentage correct data to d' did not alter any of the above findings.

Discussion. Not only were immediate recognition and long-term recognition better for the pictorial conditions than for the VD condition, but the rate of recognition loss over the retention interval was

"A SMILING OLD MAN HOLDS A LITTLE GIRL"

FIGURE 1. Example of an item in each of the 4 conditions (from top to bottom: verbal description, unembellished line drawing, embellished line drawing, and photograph).

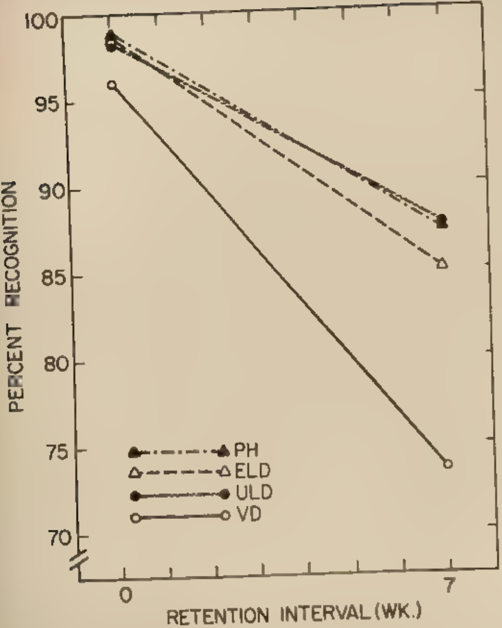


FIGURE 2. Mean percentage of correct recognitions as a function of retention interval. (Curve parameter is condition: photograph—PH, embellished line drawing—ELD, unembellished line drawing—ULD, and verbal description—VD.)

greater for VD than for the pictorial conditions. Theoretical interpretations of the latter finding in terms of forgetting should be made with caution because the high degree of original learning may or may not have been the same for the VD and pictorial conditions; however, if one is interested in the practical consequences of presenting various informational representations (verbal, pictorial) for the same amount of study time, then long-term recognition is shown to be higher for pictorial representations than for verbal representations.

The lack of effect of embellishment on pictorial recognition was quite unexpected. If the superior recognition of pictures is due to the extra details that pictures typically contain, then recognition accuracy should have occurred in the order PH > ELD > ULD. The absence of this ordering does not mean that unusual details never facilitate picture recognition, but rather that the commonplace details found in ordinary pictures do not seem to be a determinant of the superior recognition for those pictures. The theoretical problem is accounting for the lack of a facilitative effect of details while at the same time accounting for the fact (Loftus, 1972, p. 537; Poetzl, 1917) that pictorial details are entered into the information-processing system. One hypothesis that could account for both sets of findings has to do with *conservation* of processing. Suppose that for a fixed amount of processing time, *S* stores a constant amount of information about a picture, regardless of the number of details contained in the picture. That is, if one picture contains relatively few details, then *S* will store a relatively large amount of information about each detail; if another picture contains relatively many details, then *S* will store a relatively small amount of information about each detail. Because the *total* amount of stored information is the same for the 2 pictures, subsequent recognition performance would not be affected by the number of extra details (even though the extra details are stored in memory). Although this conservation hypothesis can account for the present findings about PH, ELD, and ULD, an explanation is still needed to account for the recognition advantage of pictures over verbal descriptions. The present study serves only to disconfirm the notion that this recognition advantage is due to the extra details which pictures typically contain.

REFERENCES

GELLERMAN, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *Journal of Genetic Psychology*, 1933, **42**, 206–208.

GOLDSTEIN, A. G., LOCKE, J. L., & FEHR, F. S. Children's prerecall phonetic processing of pictures and printed words. *Psychonomic Science*, 1972, **26**, 314–316.

HABER, R. N. How we remember what we see. *Scientific American*, 1970, **222**, 104–112.

LOFTUS, G. R. Eye fixations and recognition memory for pictures. *Cognitive Psychology*, 1972, **3**, 525–551.

NICKERSON, R. S. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, 1965, **19**, 155–160.

NICKERSON, R. S. A note on long-term recognition memory for pictorial material. *Psychonomic Science*, 1968, **11**, 58.

POETZL, O. Experimentell erregte Traumbilder in ihren Beziehungen zum indirekten Sehen. *Zeitschrift für Neurologie und Psychiatrie*, 1917, **37**, 278–349. Translated in *Psychological Issues*, 1960, **3**, Monograph 7, 41–120.

SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

STRONG, E. K., JR. The effect of time-interval upon recognition memory. *Psychological Review*, 1913, **20**, 339–372.

(Received March 22, 1973)

CONTENTS—(*Continued*)

9/74

Journal of Experimental Psychology

CONTENTS

ARTICLES

(*Continued on inside back cover*)

American Psychological Association

Vol. 102, No. 2 February 1974

The *Journal of Experimental Psychology* publishes original experimental investigations which are judged by the Editors to contribute substantially and significantly toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of general psychological theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Manuscripts. Submit manuscripts in *duplicate* to the Editor: David A. Grant, Department of Psychology, University of Wisconsin, Psychology Building, Charter and West Johnson, Madison, Wisconsin 53706. Manuscripts must be accompanied by an abstract of 100 to 150 words *typed on a separate sheet of paper.* The abstract should conform to the style of *Psychological Abstracts.* Detailed instructions for preparation of the abstracts appear in the *APA Publication Manual,* pp. 11–12.

The Journal offers its authors the option of a "blind" review of their submitted manuscripts under the conditions described in the Special Announcement (1972, Vol. 94, No. 3, p. 325). If the author wishes to have his identity hidden from the reviewer, he should request a blind review and take sole responsibility for preparing his manuscript in such a way as to hide his identity.

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the 1967 revision of the *Publication Manual* of the American Psychological Association. When in doubt about practices of the Journal, authors should examine a recent issue.

Clear, concise exposition is demanded; verbose manuscripts may be returned for revision before being reviewed. Piecemeal experiment-by-experiment reporting is discouraged, and integrated series experiments executed simultaneously (e.g., most doctoral dissertations) *must* be reported in a single paper. Regular papers cannot exceed 20 printed pages. Simple experiments, replications, and supplementary experiments may be published as Short Reports of not more than three printed pages. Reports of negative results and failures to replicate published work are also acceptable if they stand on their own merits as contributions to knowledge.

Normally papers are published in order of their receipt, and the current publication lag is about ten months. All of the cost of an author's alterations in galley proof is charged to the author.

Most of the editorial decisions on manuscripts submitted to the Journal are made by the Editor or one of the Associate Editors. In addition, although listed on the Masthead as Consulting Editors, Norman H. Anderson, Helen Bohmer Daly, Murray Glanzer, and Judith Goggin serve as Special Action Editors. As such, they make the editorial decisions on a limited number of manuscripts rather than serve as reviewers for the editors.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1974: $60 domestic, $61 foreign, $6 single issue. APA member rate: $25. Write to Subscription Section, APA.

Reprints: Authors may order reprints of their articles from the printer when they receive galley proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., 175 Fifth Avenue, New York, New York 10010; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. Undelivered copies resulting from address changes will not be replaced; second-class mail is not forwarded unless subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Published monthly in two volumes, six issues per volume, by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Washington, D.C., and additional mailing offices.

Journal of
Experimental Psychology

Vol. 102, No. 2 FEBRUARY 1974

LOCUS OF CONTROL AND LEARNED HELPLESSNESS [1]

DONALD S. HIROTO [2]

University of Portland

Failure to escape, the defining characteristic of learned helplessness, was investigated with perceived and instructed locus of control *S*s in a learned-helplessness paradigm. Three groups, equally divided between internals and externals and counterbalanced for sex, received different treatments with an aversive tone prior to the testing for helplessness. The first group could neither escape nor avoid an aversive tone, the second group could escape the tone, and the third group was not exposed to the treatment. Eighteen escape–avoidance trials followed, using a human analogue to an animal shuttle box, in which *S*s received an instructional set describing the task as skill or chance determined. In addition to a complete replication of learned helplessness in man, externals were significantly more helpless than internals, and chance-set *S*s were more helpless than skill-set *S*s. Since uncontrollability of noise, externality, and chance instructional set all impaired escape–avoidance in parallel ways, it was speculated that a common state may underlie all 3 dimensions—expectancy that responding and reinforcement are independent.

Overmier and Seligman (1967) and Seligman and Maier (1967) demonstrated a profound interference with shuttle box escape–avoidance behavior of dogs given prior inescapable electric shock. They hypothesized that dogs given inescapable shock failed to escape later because they had learned that shock termination was independent of responding. This learning was hypothesized to interfere with later acquisition of escape because the incentive for initiating responses had been lowered, and the ability to associate responding and shock had been proactively impaired. "Learned helplessness" was chosen as the descriptive label for the phenomenon and also as the hypothesized process by which learning of independence between responding and reinforcement interferes with future responding.

A social-learning construct, internal–external control of reinforcement (Lefcourt, 1966; Rotter, 1966), seems conceptually similar to the hypothesized learned-helplessness process. The internal–external locus of control refers to the degree to which an individual perceives that reinforcements are contingent on his actions. An "internal" tends to perceive reinforcement as a consequence of his responses and to attribute the reinforcement contingencies to his skills and abilities; an "external" tends to perceive reinforcements as un-

[1] This article is based on a dissertation submitted to the University of Portland in partial fulfillment of the requirements for the PhD degree. The author thanks the dissertation chairman, William H. James, and the members of the committee, John R. Donoghue, George S. Larimer, B. Kent Parker, and Philip Van Bruggen. Special acknowledgments go to Martin E. P. Seligman for his comments and suggestions in the preparation of this report.

[2] Requests for reprints should be sent to Donald S. Hiroto, who is now at the Counseling Center, California Polytechnic State University, San Luis Obispo, California 93401.

related, i.e., independent of his behavior, and to attribute outcomes to luck, chance, or another person.

Learned helplessness and the internal–external construct both view control of reinforcement as a crucial variable. Maier, Seligman, and Solomon (1969) and Seligman, Maier, and Solomon (1971) emphasized 2 consequences of uncontrollable events on later learning: dogs, cats, mice, goldfish, and probably rats are (*a*) slower to initiate responses to escape, and (*b*) retarded at learning that responding controls trauma. Reviews of the literature on the internal–external construct (Joe, 1971; Lefcourt, 1966, 1972; Rotter, 1966) report that external *S*s are slower than internal *S*s in learning a variety of tasks. Apparently, lack of control over reinforcement, whether real or perceived, impairs a variety of species on a variety of tasks.

A major purpose of the present study was to investigate learned helplessness in internal and external *S*s. It was predicted that *S*s given inescapable/unavoidable pretreatments would show retarded acquisition and performance measures relative to *S*s without such pretreatments, and that such retardation might be a function of externality. Previous studies with dogs as well as humans (Racinskas, 1971; Thornton & Jacobs, 1971) used electric shock as the aversive stimulus. The present study used a loud tone rather than shock as the aversive stimulus in an attempt to extend the generality of learned helplessness to a new noxious event (see also Braud, Wepmann, & Russo, 1969).

The concept of control is central to both helplessness and internal–external studies, but the definitions of the term differ. In helplessness, control refers to *E* actually arranging the events as independent of responding; while internal–external construct studies refer to perceptions of the actual events. In view of the differing reference bases of control, we introduced an instructional set to determine if perceptual set interacted with helplessness. It was predicted that *S*s pretreated with inescapable noise and then given escape–avoidance trials under an instructional set of chance would be more helpless than und[illegible]kill instructions. This parallels our pre[illegible]tion that externals would be more help[illegible]an internals following inescapable noi[illegible]

METHOD

Subjects. All *S*s were introductory [illegible]logy students at the University of Portland [illegible] administered the Internal–External form [illegible]'s Dekalb Survey Tests: Student Opinion S[illegible]y at least 8 wk. prior to the study. The Jame[illegible]ale is based on the early work of Rotter (1954) a[illegible]hares (1955) and provides a measure of the [illegible]nt to which *S* perceives reinforcements as being co[illegible]ngent on his actions or as resulting from externa[illegible]ctors. The scale is a Likert-type instrument w[illegible]4 response choices for each of the 30 critica[illegible]d 30 filler items.

A total of 96 *S*s with internal–external [illegible]res at least 1 *SD* above or below the mean were r[illegible]omly assigned to 1 of the 3 treatment groups and [illegible]of the 2 instructional-set groups, with each cell [illegible]nterbalanced for sex and locus of control.

Apparatus. Two distinctively differe[illegible] units were located on different tables. The a[illegible]ratus in the pretreatment was a red spring-loade[illegible]utton housed in a 1-in.-sq. wooden base. In the [illegible]scape condition the button was connected to rel[illegible]s controlling the termination of the aversive st[illegible]mulus, while in the inescapable condition the but[illegible]n was independent of the aversive-stimulus term[illegible]tion.

The apparatus in the phase testing for help[illegible]ness was a modified Manipulandum Type S task originally designed by Turner and Solomon (1962) as a human analogue to the 2-way shuttle box used in animal learned-helplessness studies. The manipulandum was 24 × 5 × 6 in., with a 3-in. knob protruding from the top. The knob slid on a 19 × ½ in. straight channel on the cover of the box. Attached to the knob on the underside of the channel was a 2½-in. wooden disk. The escape–avoidance response was sliding the knob to either side of the manipulandum so that the wooden disk made contact with a hidden microswitch. Only 1 of the microswitches on any trial would terminate the stimulus, so that on the next trial, the alternate switch would terminate the stimulus.

The aversive stimulus was a 3,000-Hz. tone pretaped on a Sony tape recorder and presented to *S* through North American earphones. A Beltone 15-C Audiometer recently calibrated to standards set by the International Organization for Standardization measured the stimulus at 110 db. All response variables were measured by standard (1/100-sec.) timers and counters located in an adjacent room separated by a 1-way mirror.

Procedure. Each *S* was escorted into the experimental room and informed that the study involved listening to "some loud noise which has been judged to be somewhat unpleasant but not harmful or dangerous to you." A 3-sec. sample of the 3,000-

Hz. tone was first presented, then each S was assigned to 1 of the 12 cells from the $3 \times 2 \times 2$ (Treatments $\times$ Instructional Sets $\times$ Locus of Control) factorial combination.

The treatment groups included (*a*) escape (E) *S*s, who received unavoidable/escapable pretreatments, (*b*) inescapable ($\bar{E}$) *S*s, who received unavoidable/inescapable pretreatments, and (*c*) no-pretreatment (NP) *S*s who received only the test trials with the manipulandum.

Tape-recorded instructions for pretreated *S*s were an expanded and modified version which Turner and Solomon (1962) described as adequate:

> Listen to these instructions carefully. I am not allowed to give you additional information other than what is given to you now. So please listen and do not ask me any questions. From time to time a loud tone will appear. When that tone comes on, there is something you can do to stop it.

The pretreatments consisted of 30 unsignaled 5-sec. trials with the 3,000-Hz. tone. The intertrial interval (ITI) ranged 15–25 sec., with a 20-sec. mean ITI. At the conclusion of the pretreatments, S rated the aversiveness of the auditory stimulus.

The testing for helplessness was conducted with the shuttle box manipulandum at a different table but within the same experimental room. The NP *S*s were given the preexperimental instructions and a 3-sec. sample of the tone prior to being seated at the table. The manipulandum was covered until S received 1 of the 2 instructional sets describing the task. The first half of the taped instructions was identical and were presented to all *S*s:

> You will be given some trials in which a relatively loud tone will be presented to you at different intervals. Now here is the important part, and I want you to listen carefully. Whenever you hear the tone come on there is something you can do to stop it.

The second half of the instructions varied according to the particular set. The "skill" instructions emphasized direct control over the 3,000-Hz. tone:

> What you do is really up to you to figure out. There is a solution to the problem, and if you figure it out the tone will stop. Therefore, the amount of unpleasantness you receive is dependent on your skills and abilities to find the solution to the problem. You are potentially in control of the situation.

The "chance" instructions emphasized that S had no direct control over the stimulus and that chance factors predicted success:

> But I will be controlling the solution to the problem. In other words, the way to stop the tone is really up to me. As far as you are concerned this is a guessing game. When you guess correctly, the tone will automatically stop. But if your guess is wrong the tone stays on.

After the instructions, S uncovered the manipulandum. The knob was always located at the midpoint of the manipulandum, so that S could slide the knob equidistant to either the left or right end of the box. The test phase consisted of 18 signaled 10-sec. trials. A 5-sec. red light, located at the midpoint of the manipulandum cover, preceded the onset of the 5-sec. auditory stimulus, with the offset of the light coinciding with the onset of the 3,000-Hz. tone. The ITI ranged 20–55 sec., with a mean ITI of 20 sec.

The appropriate response was moving the knob to one side of the manipulandum on one trial and sliding the knob to the opposite side on the next trial. An avoidance response was terminating the red light prior to the onset of the auditory stimulus, i.e., a response latency of 5 sec. or less, while an escape response was terminating the tone between 5 and 10 sec. after trial onset. If S did not terminate the light or noise, a latency of 10 sec. for that trial was recorded. At the completion of the test phase, S was asked to rate the unpleasantness of the tone, paid \$2, and debriefed.

Five response measures were used during the test: (*a*) trials to criterion for avoidance acquisition, defined as 3 consecutive avoidance responses; (*b*) trials to criterion for escape acquisition, defined as 3 consecutive escapes; (*c*) number of avoidance responses for the 18 trials; (*d*) number of failures to escape, defined as number of trials with a latency of 10 sec.; and (*e*) the overall mean response latency for the 18 trials. These indices, particularly *c*, *d*, and *e* parallel the indices reported in the animal helplessness literature.

RESULTS

Animal learned helplessness was characterized by the similarity in escape–avoidance behavior between E and NP *S*s, while $\bar{E}$ *S*s revealed longer response latencies and more failures to escape than either of the other *S*s. The results of this experiment disclosed remarkable similarities to the animal studies. The $\bar{E}$ group was retarded in escape–avoidance measures relative to the E and NP groups, with the latter groups not differing from each other.

Maier et al. (1969) reported that approximately 63% of $\bar{E}$ dogs and about 5% of naive dogs failed to escape in the shuttle box. Human *S*s demonstrated similar but somewhat less dramatic findings. On the average, $\bar{E}$ *S*s failed to escape the aversive stimulus on over 50% of the 18 trials, while E *S*s failed on 13% and NP *S*s failed on 11% of the trials. Approximately 34% of the $\bar{E}$ *S*s failed to reach criterion, com-

pared to 8% of the E and NP Ss who also failed to reach criterion. Figure 1 presents the response latencies of the 3 treatment groups on the 6 blocks of 3 trials. The horizontal line at the 5-sec. mark on the ordinate represents the boundary between escape and avoidance: Points plotted above the 5-sec. line represent escapes, while points below the line denote avoidance responses. The $\bar{\text{E}}$ group shows consistently longer latencies than either the E or NP groups, with the latter groups performing near equality.

An analysis of variance (ANOVA) test on response latency found a main effect of treatment, F (2, 84) = 12.38, $p < .01$, and a Treatments × Trial Blocks interaction, F (10, 420) = 2.29, $p < .05$. Two planned orthogonal comparisons between groups indicated that the main effect was due to the variability of $\bar{\text{E}}$ Ss. The first comparison between $\bar{\text{E}}$ vs. the average of E and NP Ss was significant ($p < .01$), while the second comparison found no differences between E and NP Ss. A simple main effects analysis on the Treatments × Blocks interaction disclosed significant differences between groups at each of the 6 trial blocks ($p < .02$). Scheffé S tests revealed that $\bar{\text{E}}$ Ss had consistently longer latencies at each of the trial blocks than either E or NP Ss, while the latter Ss did not differ from each other.

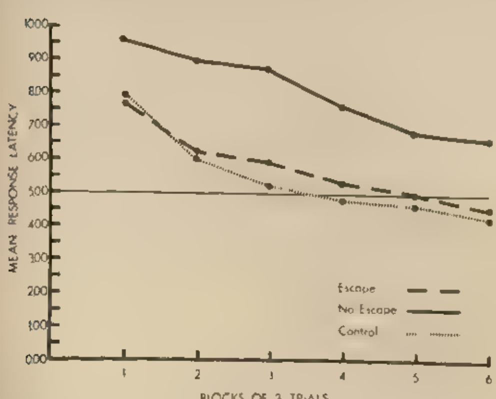


FIGURE 1. Mean response latencies of the 6 escape–avoidance trial blocks for the 3 treatment groups.

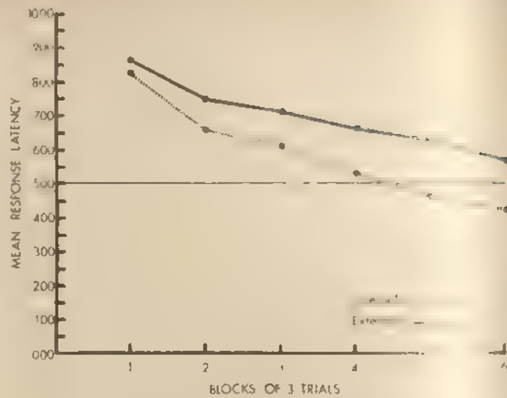


FIGURE 2. Mean response latencies of the 6 escape–avoidance trial blocks for internal and external control Ss collapsed over the 2 experimental factors.

A treatments main effect was found with the failure to escape measure, F (2, 84) = 22.57, $p < .01$, and trials to criterion for escape acquisition,[3] F (2, 66) = 1[illegible].49, $p < .01$. A planned orthogonal comparison indicated that these main effects were due to differences between E Ss and the average of E and NP Ss ($p < .01$). This planned comparison also found a significant effect on the trials to criterion for avoidance acquisition: $\bar{\text{E}}$ Ss took longer to reach acquisition than the average of E and NP Ss, F (1, 84) = 4.45, $p < .05$. The second orthogonal comparison found no differences between E and NP Ss on any of the 4 dependent measures.

Other factors, independent of treatment, were also significant. External locus of control Ss, regardless of their pretreatments and instructional sets, were slower to escape or avoid than internal control Ss. Figure 2 presents the response latencies of the internal–external factor on the 6 trial blocks. An ANOVA on response latencies disclosed an internal–external main effect, F (1, 84) = 6.58, $p < .05$, and a Trial Blocks × Internal–External interaction, F (5, 420) = 3.24, $p < .05$.

[3] Analysis of the escape-acquisition measure was based on unequal ns as some Ss (18) reached avoidance criterion without first reaching escape criterion. There was no statistical relationship between the 18 Ss dropped from the analysis and any of the experimental treatment factors.

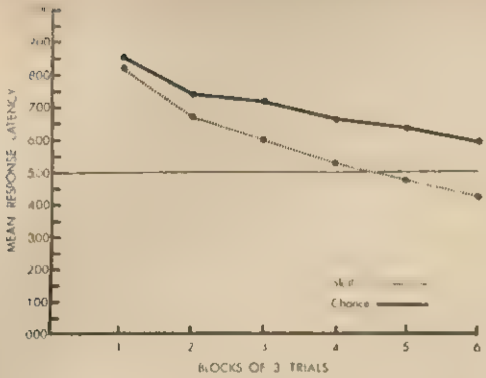


FIGURE 3. Mean response latencies of the 6 escape–avoidance trial blocks for the instructional-set *S*s collapsed over the 2 experimental factors.

A test for simple main effects on the interaction revealed that external control *S*s responded slower than internal control *S*s on Trial Blocks 4, 5, and 6 ($p < .025$). External *S*s also required more trials to reach avoidance criterion ($M = 18.3$) and made fewer avoidance responses during the 18 trials ($M = 3.1$) than internal *S*s (*M*s = 15.0 and 6.2, respectively). The ANOVA tests on these measures statistically verified the observed differences ($p < .05$). There were no internal–external differences for the escape-acquisition or failures-to-escape measures.

The chance-set group displayed retarded measures compared with the skill-set group. Figure 3 presents the response latencies of the instructional set factor collapsed over the 2 experimental factors. The ANOVA found a main effect of instructions, $F(1, 84) = 5.93$, $p < .05$, and a Trial Blocks × Instructions interaction, $F(5, 420) = 4.28$, $p < .05$. A simple main effects analysis indicated that chance-set *S*s had longer latencies than skill-set *S*s on Trial Blocks 4, 5, and 6 ($p < .025$). The effects of instructional set also interfered with avoidance responding. Compared to skill-set *S*s, chance-set *S*s required more trials to reach avoidance criterion (*M*s = 18.2 and 15.1, respectively) and made fewer avoidance responses (*M*s = 3.2 and 6.1, respectively). The ANOVAs computed on these measures were all significant ($p < .05$). There was no effect of instructions on either the escape-acquisition or failures-to-escape measure.

The results of internal vs. external and instructional set indicate that either factor can produce parallel results of inescapability. That is, inescapability, chance set, and externality retard escape–avoidance behavior.

Interestingly $\bar{E}$ internal *S*s made more button presses i.e., escape attempts, in pretreatment than $\bar{E}$ external *S*s (*Mdn*s = 21 and 5.5, respectively), $\chi^2 = 3.97$, $p < .05$, suggesting that persistence in trying to control an uncontrollable event may be correlated with the absence of later helplessness. This is similar to Seligman and Maier's (1967) concept of immunization where dogs with prior escapable shock demonstrated enhanced panel pressing under inescapable shock and were later not helpless.

There were no significant interactions between the experimental factors on any response measure. However, a priori comparisons permitted an analysis of possible interaction within the E group. Figure 4 presents the mean latencies of internal and external $\bar{E}$ *S*s across the 6 trial blocks. The *F* test for planned orthogonal comparisons revealed that the overall mean latency of externals was statistically longer than that of internals, $F(1, 84) = 5.46$, $p < .05$. No other dependent measure approached significance. Inescapability and instructional set were

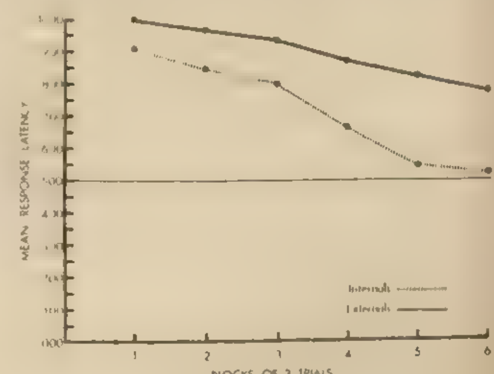


FIGURE 4. Mean response latencies of internal and external control *S*s in the inescapable treatment group for the 6 escape–avoidance trial blocks.

analyzed next. Dunn's test for planned nonorthogonal comparisons was not significant with any of the 5 response measures.

Further analysis was conducted on a possible confounding variable. There was no yoking procedure in the pretreatment, so that duration of noise exposure was not equated. The Ē group received 5 sec. of aversive noise on each trial, whereas the E group received an average of 1.4 sec. of noise per trial. Since Ē Ss received longer periods of the loud tone, and perhaps greater stress, it could be argued that differential stress rather than uncontrollability led to the retarded performance with the shuttle box manipulandum. However, *S* ratings of the noise did not reveal significant differences: the Ē group rated the 110-db. tone at 3.55 on a 7-point scale, while E group rated the same tone at 3.66. This difference was not statistically different and suggests that differential stress did not produce the helplessness effects. It should also be noted that the relatively moderate ratings of the 110-db. stimulus indicates that learned helplessness can be produced with moderately aversive events as well as the more traumatic events used in the animal studies.

Discussion

Learned helplessness can be experimentally produced in man. Both animals and man show longer latencies and more failures to escape following inescapable aversive events than following escapable events or no pretreatment. The inescapable pretreatments did not affect the total number of avoidance responses between groups (see also Seligman & Maier, 1967). It should be noted that Overmier (1968) reported interference with avoidance acquisition when escape contingencies were eliminated in the shuttle box. Considering the divergent findings, perhaps it is only initial learning—escape or avoidance, whichever occurs first—that is disrupted by inescapability.

The first published account of learned helplessness in humans (Thornton & Jacobs, 1971) had 2 methodological problems. First, an instructional set was confounded with inescapability in the pretreatment. The Ē *Ss* received instructions that described the noncontingency between shock and responding, and E *Ss* received different instructions that described the contingency between shock and avoidance. These different instructions, embedded within the different pretreatments, may have predetermined *Ss*' responses to [illegible] a relationship in the study. The second problem relates to the unusual pretreatment procedures. Previous animal helplessness studies pretreated Ē *Ss* with unavoidable-inescapable shock. Thornton and Jacobs, however, pretreated their Ē *Ss* with avoidable-escapable shock. Because of the different procedures, their study cannot be considered an analogue to the helplessness studies. Thornton and Jacobs, rather than demonstrating learned helplessness, demonstrated the effect of prior avoidance training on a later escape-avoidance task.

The learned-helplessness hypothesis identifies control over reinforcement as the crucial variable. The present study confirmed the hypothesis by demonstrating that Ē Ss produced greater impairment in escape-avoidance than E or NP *Ss*. In addition, the external locus of control variable, also concerned with control of reinforcement, interacted with inescapability to produce greater impairment than internal control *Ss*.

In conclusion, this study demonstrated that helplessness can be experimentally induced in man wholly parallel to helplessness in animals. The *S* variable of externality appears to function like the pretreatment variable of inescapability, as evidenced by the interaction between externality and the helplessness induction. Both locus of control and skill-chance instructional set factors produced an effect similar to inescapability. In view of the parallel effects between the 3 factors, I suggest that a single process may underlie learned helplessness, externality, and the perceptual set of chance—the expectancy that responding and reinforcement are independent. Seligman (in press) and Miller and Seligman (1973) have pointed to this cognition as the underlying state in reactive depression in man.

References

Braud, W., Wepmann, B., & Russo, D. Task and species generality of the "helplessness" phenomenon. *Psychonomic Science*, 1969, **16**, 154-155.

Joe, V. C. Review of the internal-external control construct as a personality variable. *Psychological Reports*, 1971, **28**, 619–640.

Lefcourt, H. M. Internal versus external control of reinforcement: A review. *Psychological Bulletin*, 1966, **65**, 206–220.

Lefcourt, H. M. Internal versus external control of reinforcement revisited: Recent developments. In B. A. Maher (Ed.), *Progress in experimental personality research.* Vol. 6. New York: Academic Press, 1972.

Maier, S. F., Seligman, M. E. P., & Solomon, R. L. Pavlovian fear conditioning and learned helplessness: Effects on escape-avoidance behavior of a) the CS-UCS contingency and b) the independency of the US and voluntary responding. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

Miller, W. R., & Seligman, M. E. P. Depression and the perception of reinforcement. *Journal of Abnormal Psychology,* 1973, **82**, 62–73.

Overmier, J. B. Interference with avoidance behavior: Failure to avoid traumatic shock. *Journal of Experimental Psychology,* 1968, **78**, 340–343.

Overmier, J. B., & Seligman, M. E. P. Effects of inescapable shock upon subsequent escape and avoidance responding. *Journal of Comparative and Physiological Psychology,* 1967, **63**, 28–33.

Phares, E. J. Changes in expectancy in skill and chance situations. Unpublished doctoral dissertation, Ohio State University, 1955.

Racinskas, J. R. Maladaptive consequences of loss or lack of control over aversive events. Unpublished doctoral dissertation, University of Waterloo, 1971.

Rotter, J. B. *Social learning and clinical psychology.* Englewood Cliffs, N.J.: Prentice-Hall, 1954.

Rotter, J. B. Generalized expectancies of internal versus external control of reinforcements. *Psychological Monographs,* 1966, **80** (1, Whole No. 609).

Seligman, M. E. P. Depression and learned helplessness. In R. J. Friedman & M. M. Katz (Eds.), *The psychology of depression: Contemporary theory and research,* in press.

Seligman, M. E. P., & Maier, S. F. Failure to escape traumatic shock. *Journal of Experimental Psychology,* 1967, **74**, 1–9.

Seligman, M. E. P., Maier, S. F., & Solomon, R. L. Unpredictable and uncontrollable aversive events. In F. R. Brush (Ed.), *Aversive conditioning and learning.* New York: Academic Press, 1971.

Thornton, J. W., & Jacobs, P. D. Learned helplessness in human subjects. *Journal of Experimental Psychology,* 1971, **87**, 367–372.

Turner, L., & Solomon, R. L. Human traumatic avoidance learning: Theory and experiments on the operant-respondent distinction and failures to learn. *Psychological Monographs,* 1962, **76** (40, Whole No. 559).

(Received March 5, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 194–198

CONDITIONED ADAPTATION TO PRISMATIC DISPLACEMENT: TRAINING TRIALS AND DECAY[1]

JEROME H. KRAVITZ[2]
Howard University

FRED L. YAFFE
Eye Research Foundation, Bethesda

Adaptation to prismatic displacement was conditioned to the wearing of a pair of goggles by employing Taylor's alternation technique. Each alternation consisted of 2 3-min. training conditions. In one condition, *S* was exposed to the goggles and the prism; in the other, *S* was exposed to neither. Three groups of 13 *S*s were trained. They received either 1, 3, or 5 alternations. Pointing test measures designed to measure aftereffects and conditioned aftereffects were performed by each *S* at 0, 5, 10, and 20 min. after training. Otherwise, *S* sat quietly with closed eyes after training. Training time and testing time both produced significant effects. Aftereffects of adaptation exhibited a reminiscence effect and then decay, whereas the conditioned aftereffects exhibited neither.

When a prism displaces the viewer's visual field it initially causes errors in localization responses (effect). These errors decrease in magnitude (adaptation effect) after some appropriate practice by *S*. The adaptation effect measures are in turn taken to reflect an underlying psychological compensation (adaptation) to the optical disarrangement. Responses then made with the removal of the device exhibit an error in the direction opposite to the effect (aftereffect). For some time, aftereffects were assumed to be conservative estimates of adaptation effects and as an equally valid indicator of adaptation because, in the early studies which measured both, the magnitude of the former so nearly agreed with the magnitude of the latter (Rock, 1966). Conditioned adaptation studies, however, have indicated that this equality need not always hold (Kohler, 1964).

Conditioned adaptation studies usually have 2 major parts: a training period and testing periods. The training period is designed to produce an association between the adaptation and a previously neutral stimulus. This is achieved (*a*) by employing 2 alternating training conditions, one with normal vision and the other with vision modified by the optical device, and (*b*) by the systematic covariation with the training exposure conditions of other different processes, specifically the absence and presence of the neutral stimulus.

To date, 2 techniques have been developed to produce this alternation. In one (Kohler, 1964), *S* is exposed to a split visual field by having an optical device, a prism, in the upper half of the field and unobstructed normal viewing in the lower. Here the alternation produced is solely dependent upon a behavior of *S*, namely *S*'s changing the direction of his gaze. In the other technique (Taylor, 1962), the alternation of the training conditions depends upon the intervention of *E*, who simultaneously exposes *S* to the optical device and the neutral stimulus and later removes them. Conditioned adaptation effects are taken to be present when in the testing period the adaptation effects or aftereffects obtained with the neutral stimulus present are greater than those obtained with it absent.

Taylor (1962) produced differential conditioned adaptations by systematically covarying the training exposure conditions with the sensations produced by wearing spectacles that housed the optical device. With this procedure he conditioned the adaptation effect to right–left reversal and to the prismatically produced slant of horizontal surfaces. During testing he found that the adaptation effects obtained while wearing the spectacles were much larger than the aftereffects obtained with the spectacles absent.

---

[1] This study was supported by National Institute of Neurological Disease and Stroke Grant NB 08696-01.

[2] Requests for reprints should be sent to Jerome H. Kravitz, Department of Psychology, Howard University, Washington, D.C. 20001

Recently Kravitz (1972) refined the Taylor (1962) alternation training technique and associated the adaptation to prismatic displacement to the wearing of goggles in 240 min. of training. This study replicated Taylor's findings, namely, that there were larger adaptation effects with the goggles in place than aftereffects with them removed. It also demonstrated that aftereffects obtained while wearing the goggles without a prism were larger than those obtained without goggles. The magnitudes of these differences were taken as a measure of the conditioned adaptation to the goggles. Aftereffects obtained in the absence of the goggles are accordingly viewed as measuring a component of the total adaptation not under the control of the goggles as a conditional stimulus (CS). One current theoretical issue regards how adaptation in general and this component in particular should be viewed. However, for reference purposes we propose to provisionally label this component as "generalized aftereffect." One position holds that adaptation does not depend upon or necessarily result in learning (Mikaelian, 1963). The other holds that adaptation not only depends upon past experience (Held, 1961) but is due to learning (conditioning) processes (Taub, 1968). If the generalized aftereffect is a form of learning, then it could be expected to "decay" (Ebenholtz, 1969; Hamilton & Bossom, 1964) in a manner similar to that of conditioned adaptations. The present study tests this hypothesis.

Method

Apparatus. The displacement of the visual field was produced by a 20-diopter wedge prism mounted, base left, in the right frame of a pair of welder's goggles. The left eye of *S* was occluded at all times by an eye patch. A second pair of welder's goggles had an optically plain piece of glass in its right frame.

The testing apparatus was a vertically adjustable, horizontal platform, 152 cm. long and 66 cm. deep, which when raised to shoulder level allowed *S* to extend his arm easily underneath it. On the platform's near edge a biteboard apparatus was fixed. On the far side of the platform was a panel (102 cm. wide by 76 cm. high) perpendicular to *S*'s line of regard. The pointing test target was a vertically oriented black bar (6 × .6 cm.). It was presented in 2 positions: 6° to the right of *S*'s midline and 6° to the left, through a 76-cm.-long, horizontal, eye-level slit in the rear panel. The rear panel and the remainder of the apparatus were painted flat white. The *E* used a meter stick to determine *S*'s pointing responses.

Responses. A manual localization test was employed to measure the effects of training. It involved having *S* stand biting the biteboard, extending his right arm under the platform, and pointing with his right index finger. At the testing apparatus, *S* had his eyes closed at all times, except when making a pointing response. The pointing test measure was the mean of 4 pointing responses: 2 responses to each of 2 target positions. Each response required *S* to open his eyes and point at the apparent location of the target. The order of target position presentation was counterbalanced.

Test conditions. The pointing test was performed in 2 testing situations: In the first $(G/\bar{P})$, *S* wore the goggles (G) with plain glass $(\bar{P})$; in the second $(\bar{G}/\bar{P})$, *S* did not wear the goggles $(\bar{G})$ or the prism $(\bar{P})$.

Experimental design. Thirty-nine undergraduates drawn from Howard University introductory psychology courses served as paid voluntary *Ss*. They were assigned to 3 equal groups.

Each group was administered the same 6 main segments of the experiment, which were in order of occurrence: pretesting, training, and 4 posttest segments (A, B, C, and D). Figure 1 depicts the manner in which the 2 types of test situations were distributed among the first 3 segments of the experiment. The subscripted numbers affixed to each test condition indicate its order of occurrence. Each *S* went through the following sequence. First was the pretesting, which involved 2 test conditions: $(\bar{G}/\bar{P})_1$ and $(G/\bar{P})_2$. The *S* then received alternation training followed by Posttesting A. All the posttest segments had the same test condition sequence. The Posttest A sequence had 3 test conditions: $(\bar{G}/\bar{P})_3$, $(G/\bar{P})_4$, and $(\bar{G}/\bar{P})_5$. Posttest B immediately followed Posttest A. After Posttest B there was a 5-min. rest period followed by Posttest C. Posttest C was followed by a 10-min. rest period and then Posttest D. Thus, Posttests A, B, C, and D occurred 0, 5, 10, and 20 min. after the completion of training. During each rest period *S* sat quietly with eyes closed.

The effects of training and/or decay were determined by obtaining the difference between the pre- and posttesting measures of a given type of test condition.

Since all testing was under $\bar{P}$ (no prism) conditions, these differences (represented in Figure 1 by solid lines connecting the pretest conditions to the Posttest A conditions) resulted in aftereffect measures represented in Figure 1 by the circled d symbol appended to each solid line.

Thus, we obtained from each posttest segment 3 pointing test measures of aftereffects of adaptation. It was expected that the G aftereffects would be greater than the $\bar{G}$ aftereffects to the extent that the adaptation had become conditionally associated to the wearing of the goggles. The difference between these effects (indicated in Figure 1 as the

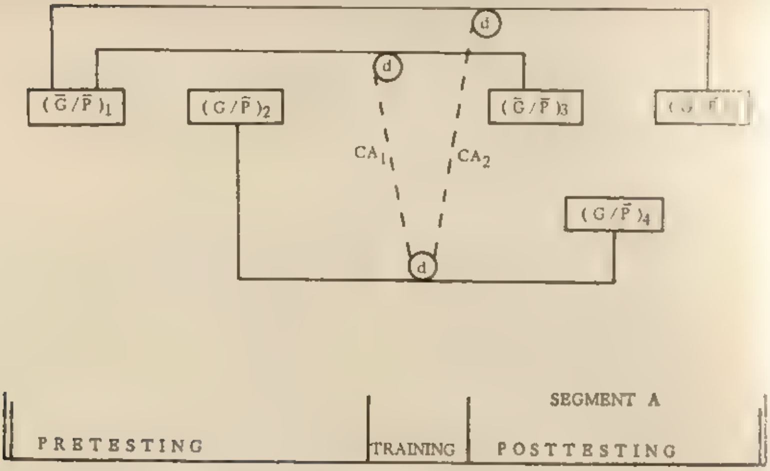


FIGURE 1. The pretest, training, and posttest (Segment A) sequence. (The solid lines represent differences between pre- and posttest conditions that result in the measurement of aftereffects [circled d] of adaptation. Dashed lines connecting circled d symbols represent differences between G [goggles present] and Ḡ [goggles absent] effects and result in the measurement of conditioned adaptation effects [CA_1, CA_2].)

dashed lines labeled CA_1 and CA_2 connecting the circled d symbols, Posttest A) are considered measures of conditioned adaptation (CA). The mean of these 2 measures (CA_e) constitutes a single estimate of the conditioned effects measured in the testing segment. The same calculations apply to all of the testing segments.

Training procedure. Each group (E_1, E_3, and E_5) was given alternation training designed to produce differential conditioned adaptation to the wearing of goggles. Group E_1 received 1 alternation; E_3, 3; and E_5, 5. The total training times of E_1, E_3, and E_5 were 6, 18, and 30 min., respectively. Each alternation consisted of 2 3-min. training exposures: In G/P, *S* wore the goggles (G) and was exposed to prismatically altered viewing (P); in Ḡ/P̄ the goggles were absent (Ḡ) and the viewing was normal and unaltered (P̄). The alternation was accomplished by having *S* sit with his eyes closed while *E* placed or removed the goggles. During each training exposure, whether G/P or Ḡ/P̄, *S* performed the same training task. The task consisted of viewing his right arm resting upon a vibrating armrest. Twice during each 3-min. segment, *S* was instructed to move his arm and hand to the left, center, right, and center of the arm rest. Otherwise, the arm rested passively (Kravitz & Wallach, 1966). Each alternation began with a G/P exposure and ended with a Ḡ/P̄ exposure.

RESULTS

The principal inquiries of this study were: (*a*) Will conditioned adaptation effects be produced by short forms of training; and (b) what is the relationship between the decay of generalized aftereffects and the decay of conditioned adaptation effects? Therefore, only Ḡ aftereffects and CA_e effects are presented in Table 1. Each Ḡ figure is the mean difference in degrees between the mean of the 2 Ḡ/P̄ posttest conditions of a segment and the corresponding pretest condition. The CA_e effects are a derived measure. They are the mean differences between the Ḡ and the G aftereffect measures. Each G measure (not presented in the table) is the mean difference between the pre- and posttraining measurements of the G/P̄ test conditions.

Individual *t* tests were performed on each of these mean effects to determine whether they were significantly different from zero. Inspection of Table 1 indicates that even with the short training times employed, significant conditioned adaptation effects were obtained. The grand mean CA_e effect, 1.5°, was significant, $t(38) = 4.43$, $p < .01$.

A 3-way (3 × 4 × 2) analysis of variance with 2 repeated factors was performed and is presented in Table 2 (Winer, 1962). The first factor was training groups (6, 18, and 30 min.); the second factor was testing time (0, 5, 10, and 20 min.); and the third

TABLE 1

Mean $\bar{G}$ Aftereffects and Mean CA_c Effects (in degrees) on Pointing Test

Training time (n = 13/group) (df = 12) Total df = 38)	Posttesting segment: A (0 min.)		B (5 min.)		C (10 min.)		D (20 min.)		M	
	$\bar{G}$	CA_c	$\bar{G}$	CA_c	$\bar{G}$	CA_c	$\bar{G}$	CA_c	$\bar{G}$	CA_c
6 min.										
$\bar{X}$	.0	1.5**	.0	1.0	− .9	.9*	−1.7*	.0	− .6	.8*
SD	2.1	1.7	2.4	2.3	2.6	1.8	2.1	2.1	2.1	1.7
18 min.										
$\bar{X}$	1.0*	1.7**	2.1**	1.0**	1.4*	.7	.2	1.0*	1.2*	1.1**
SD	2.4	2.2	2.6	1.3	2.8	1.9	2.0	1.8	2.6	1.4
30 min.										
$\bar{X}$	1.4*	2.4**	2.6**	2.4**	1.1*	2.5**	.8	2.6**	1.5**	2.5**
SD	2.1	1.7	2.6	1.9	2.8	2.1	4.6	1.7	2.3	1.5
Total										
M	.8**	1.9**	1.5*	1.5**	.5	1.4**	− .2	1.2**	.7*	1.5**
SD	2.3	1.6	2.9	1.9	2.9	2.1	3.2	2.1	2.5	1.7

Note. Each $\bar{G}$ is the mean difference between the mean of the 2 $\bar{G}$ $\bar{P}$ (goggles and prism absent) posttest conditions of a segment and the corresponding pretest conditions. A minus signifies a $\bar{G}$ posttest deviation in the nonadaptive direction.) Each CA_c is the mean difference between the $\bar{G}$ and the G aftereffect measures. (The G measure is the mean difference between the pre- and posttraining measures of the $\bar{G}/P$ test conditions.)
* $p < .05$; deviations significantly different from 0; 2-tailed test.
** $p < .01$; deviations significantly different from 0; 2-tailed test.

factor was the comparison of aftereffects and conditioned aftereffects ($CA/\bar{G}$). The group effect was significant, F (2, 36) = 8.05, $p < .01$. A Newman-Keuls comparison of the group ordered means revealed that the 30-min. group produced significantly larger effects than both the 18- and 6-min. groups ($p < .05$), and that the 18-min. group was correspondingly larger than the 6-min. group ($p < .05$).

The time effect was also significant, F (3, 108) = 11.51, $p < .01$. A Newman-Keuls comparison of the time effects revealed that the immediate test was not significantly different than the 5-min. test. All other comparisons revealed significant decrements in effects with time ($p < .01$). However, these decrements are accounted for by the Time × $CA/\bar{G}$ interaction, which was the only other significant effect of the analysis, F (3, 108) = 4.02, $p < .01$. Although the CA_c trend is decreasing, a Newman-Keuls comparison of this component of the interaction revealed that these effects did not significantly change with time. The same comparison of the $\bar{G}$ effects (generalized aftereffects obtained with the CS absent) revealed significant changes as a function of time. Apparently the CA_c decay rate is slower than that of generalized aftereffects. The comparison of the immediate $\bar{G}$ effect to the 5-min. effect revealed a "reminiscence effect": The 5-min. effect was significantly larger (.7°) than the immediate effect ($p < .05$). The $\bar{G}$ effects then progressively decreased in magnitude with time. The 10- and 20-min. effects were both significantly smaller (1.0° and 1.7°, respectively) than the 5-min. effect ($p < .05$). The 20-min. effect was also smaller (1.0°) than the immediate effect ($p < .05$).

TABLE 2

Summary of 3-Factor Analysis of Variance

Source	df	MS	F
Between	38		
Groups (A)	2	2,296.82	8.06*
Error	36	285.08	
Within	273		
Time (B)	3	420.55	11.51*
A × B	6	58.85	1.61
Aftereffects vs. conditioned aftereffects (C)	1	1,090.88	2.54
A × C	2	464.70	1.08
Error	36	428.66	
B × C	3	192.10	4.02*
A × B × C	6	29.26	.61
Error	108	47.75	

* $p < .01$.

Thus, an increase in the number of trials and training time does affect the magnitude of the conditioned adaptation effects and aftereffects obtained. Conditioned adaptation effects do not decrease in magnitude due to passive forgetting; generalized aftereffects obtained in the absence of the CS reveal at first an increment and then a progressive decrement with time.

Discussion

The CA_c effects of the present study indicate that relatively short periods of training can produce conditioned adaptation effects. An argument against this conclusion might be that these effects are due to some form of conscious correction by *S*. Such a correction would involve implicit verbal self-instruction regarding the effects of the prism, its contingencies, and a mode of corrective compensatory responding.

Evidence, however, does not support this contention. First, in postexperimental interviews most *S*s were unable to verbalize either what happened in the experiment or the displacement effects of the prism. All denied deliberate compensation during testing.

Second, the magnitude of the overall CA_c effects (1.5°) is so small relative to the prismatic displacement (11.6°) that an explanation of these effects in terms of conscious correction would constitute a gross extension of the meaning of that construct. We therefore conclude that the CA_c effects are not due to conscious correction and that they indicate that conditioning has been produced.

Aftereffects typically increase in magnitude with increasing training time (Hamilton, 1964). The results of the present study indicate that this relationship holds even under conditions of alternation training. In addition, we find that conditioned adaptation effects and generalized aftereffects both increase with increasing amounts of training and in this regard function similarly. Ebenholtz (1969) found that tilt aftereffects decay with time and that the rate of decrease was independent of initial magnitudes of effects. The lack of a Groups × Time interaction in the present study extends his finding to aftereffects obtained under alternation training. Ebenholtz also developed the comparator model of adaptation (Held, 1961) so as to account for the obtained decay of tilt aftereffects. However, that model, as presently formulated, cannot account for the reminiscence effect obtained with the $\bar{G}$ aftereffects in th[illegible] study. Moreover, the present effect repli[illegible] both in terms of time and direction that o[illegible] with monkeys (Taub, 1968). This [illegible] suggests that generalized aftereffects and [illegible] ad- aptation effects may be affected [illegible] more or less transient inhibitory proce[illegible] Other- wise, the decay of $\bar{G}$ aftereffects w[illegible] as expected. The adaptation as a form of learning hypothesis would lead to the deduction that generalized aftereffects and co[illegible] after- effects would decay similarly. Ou[illegible] lts do not confirm this deduction. Ra[illegible] they imply that the hypothesis must [illegible] dified if it is to remain viable. The n[illegible] ation would be such as to take into accou[illegible] fact that adaptation as a form of lea[illegible] and conditioned adaptation as a form o[illegible] rning function dissimilarly in some repects.

References

Ebenholtz, S. Transfer and decay f[illegible] ons in adaptation to optical tilt. *Journal o[illegible] xperimental Psychology*, 1969, **81**, 170–173.

Hamilton, C. R. Intermanual transfer o[illegible] laptation to prisms. *American Journal of Psychology*, 1964, **77**, 457-462.

Hamilton, C. R., & Bossom, J. Decay of prism aftereffects. *Journal of Experimental Psychology*, 1964, **67**, 148–150.

Held, R. Exposure-history as a factor in maintaining stability of perception and coordination. *Journal of Nervous and Mental Diseases*, 1961, **132**, 26–32.

Kohler, I. The formation and transformation of the perceptual world. *Psychological Issues*, 1964, **3** (4, Monograph No. 12).

Kravitz, J. H. Conditioned adaptation to prismatic displacement. *Perception & Psychophysics*, 1972, **11**, 38–42.

Kravitz, J. H., & Wallach, H. Adaptation to displaced vision contingent upon vibrating stimulation. *Psychonomic Science*, 1966, **6**, 465–466.

Mikaelian, H. H. Failure of bilateral transfer in modified hand–eye coordination. Paper presented at the meeting of the Eastern Psychological Association, New York, April 1963.

Rock, I. *The nature of perceptual adaptation.* New York: Basic Books, 1966.

Taub, E. Prism compensation as a learning phenomena: A phylogenetic perspective. In S. J. Freedman (Ed.), *The neurophysiology of spatially oriented behavior.* Homewood, Ill.: Dorsey Press, 1968.

Taylor, J. G. *The behavioral basis of perception.* New Haven: Yale University Press, 1962.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received October 20, 1972)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 199–208

PERCEPTUAL UNITS IN SPEECH RECOGNITION [1]

DOMINIC W. MASSARO [2]

University of Wisconsin—Madison

The size of the sound stimulus employed in the first stage of speech processing was investigated in an attempt to determine the perceptual unit of analysis in speech recognition. It is assumed that the perceptual unit is held in a preperceptual auditory image until its sound pattern is complete and recognition has occurred. Vowels and consonant–vowel syllables were employed as test items in a recognition-masking task. The results show that recognition performance improved up to 200–250 msec. after presentation of the speech sound. The results were interpreted as evidence that the preperceptual auditory storage and perceptual processing of a speech sound does not exceed 250 msec., implying that some transformation of the speech signal must occur about every ¼ sec. Since the stimulus within this time period must function as a perceptual unit, perceptual units appear to be of roughly syllabic length.

The primary purpose of this study was to determine the size of the sound stimulus employed in the first stage of speech recognition. A listener has recognized (identified) a stimulus when he has determined that one of a possible set of alternatives was presented. Recognition of a stimulus is possible only if the information in the stimulus is sufficient to distinguish that stimulus from other possible stimulus alternatives. Recognizing speech continuously implies that each small portion of the sound-wave pattern uniquely determines a stimulus alternative. However, small portions of the acoustic input are not unique; there is no one-to-one mapping of stimulus to percept. Since the sound pattern must contain enough information for a consistent stimulus–percept mapping, larger chunks of the acoustic input are necessary for recognition.

Since the sound stimulus for speech recognition extends over time, the first part of the pattern must be held in some preperceptual form until the pattern is complete and recognition has occurred. Accordingly, it is assumed that the information in the sound pattern is held in a preperceptual auditory image and that the recognition process involves a readout of the information in that image. The duration of the preperceptual auditory image places an upper limit on the sound patterns that can be employed in the recognition process. Accordingly, to understand speech recognition, it is necessary to determine the maximum duration that information can be held in a preperceptual auditory image.

The minimal sound patterns that are usually recognized in continuous speech are referred to as perceptual units. In terms of the present analysis, the perceptual unit cannot exceed the duration of the image and must uniquely determine the appropriate perceptual response. The perceptual unit of analysis gives the perceptual system acoustic information that can be reliably correlated with information in long-term memory. The information in the unit is defined by a set of features that correspond to a list of features in long-term memory. The features made available by the stimulus are acoustic, whereas the list of features in memory is abstract. The recognition process must find a match between the acoustic features in the stimulus and a list of features in memory. The nature of perceptual units also provides information about the structure of long-term memory. A set of features in the acoustic signal indicates that the signal can be identified and can therefore be used to

[1] This investigation was supported in part by U.S. Public Health Service Grant MH-19399-02. A report of this research was presented at the Indiana Conference on Theoretical and Mathematical Psychology, Bloomington, April 1972. The reviewer provided a number of helpful comments on the manuscript.

[2] Requests for reprints should be sent to Dominic W. Massaro, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

integrate with other units so that meaning can be derived from the message.

The following studies were carried out to determine the size of perceptual units employed in speech perception. It is necessary, therefore, to determine the temporal course of processing speech stimuli and to estimate the effective duration of the preperceptual image. Vowels and consonant–vowel syllables were employed as test items in an auditory recognition masking task (Massaro, 1970b). In this task, a test stimulus is presented, followed by a masking stimulus that occurs after a variable silent interstimulus interval. In a typical experiment, the test signals are short tones differing in pitch, and the listener's task is to identify the higher tone as high and the lower tone as low. The masking tone is presented at another frequency and at the same loudness as the test tones. The results indicate that recognition performance improves with increases in the silent intertone interval, up to about 250 msec. (Massaro, 1970b, 1971, 1972c).

These results provide information about the preperceptual auditory image of the test tone and the vulnerability of the auditory image to new inputs. Given that the test tone lasted only 20 msec., some preperceptual image must have remained for the perceptual processing necessary to improve recognition performance with increases in the silent intertone interval. This improvement in recognition also indicates that the masking tone terminated perceptual processing of the image. Since recognition performance leveled off at about 250 msec., the image effectively decayed within this period. Using this paradigm with speech stimuli should make it possible to determine the duration of preperceptual images and the processing time required for speech stimuli.

Experiment I

The first experiment demonstrates that the recognition of a short speech stimulus is not immediate, but rather requires time for perceptual processing. Perceptual processing refers to the analysis of physical features in the sensory input in order to recognize the stimulus. A measure of perceptual-processing time might be found in the durations of vowels in normal speech which are in the range of 150 [illegible] msec. (Fletcher, 1953; House, 1961; Peterson & Lehiste, 1960). Since the pattern of sound-pressure fluctuations in a steady-state vowel repeats at the speaker's fundamental frequency, the extended duration of the vowel might be necessary for processing the information available in the vowel presentation. Vowels presented for very short durations can be identified if followed by a silent retroactive interval (Gray, 1942; Suen & Beddoes, 1972). However, if processing is interfered with by following the short test-vowel presentation with another speech sound, the test vowel should not be identified. This result would provide evidence that the duration of the vowel in normal speech allows time for processing, since the extended duration of the vowel protects it from later speech until recognition has been completed.

The *S*s were required to recognize a test vowel. On each trial, 1 of 2 vowels could be presented with equal probability. A second vowel, referred to as the masking vowel, followed the test vowel after a variable silent intervowel interval. It was assumed that the perceptual processing necessary for recognition of the test vowel could take place during the test-vowel presentation and the silent interval afterwards, but *not* during the masking-vowel presentation.

Method

Subjects. Three young adults from the University of Wisconsin community were employed as *S*s.

Procedure. The vowels of a male speaker were first recorded at the same fundamental frequency and amplitude. A steady-state segment of each vowel was stored digitally by a computer-controlled analog-to-digital converter. During this experiment, the vowel segment was played back, using a digital-to-analog converter. In the recognition masking task, the vowels /i/ as in *heat* and /I/ as in *hit* were employed as test items. The duration of the test vowel was 20 msec. The silent intervowel interval lasted 0, 20, 40, 80, 160, 250, 350, or 500 msec. The masking stimulus was a 270-msec. nonsense vowel made up of 2 alternating vowel segments. These 2 segments were taken from the

vowels /a/ as in *hat* and /U/as in *put* and lasted 45 msec. each.

The *Ss* were tested simultaneously in a sound-insulated chamber. All experimental events were controlled by a PDP-8 computer. The vowels were presented binaurally over matched headphones (Grason-Stadler Model TDH-39) at a normal listening intensity (about 80-db. SPL). An *S* recorded his decision by pressing 1 of 2 push buttons, labeled i and I, respectively. Following the 2.5-sec. response period, feedback was given by illuminating a small light for 500 msec. above the correct response button. The intertrial interval was 2 sec.

On every trial, *S* heard a test vowel, followed in turn by a variable silent interval and the masking vowel. He identified the test vowel as i or I and was then informed of the correct answer for that trial. The *Ss* had 1 day of practice identifying the test vowels without a masking vowel and 2 days of practice with a masking vowel before the experiment itself, which consisted of 2 days of 3 sessions each. There were 400 trials/session, and *Ss* did not respond to the first 5 trials of each session. Since all experimental conditions were completely random, memory and decision factors were reasonably constant. The results, then, indicate the temporal course of the perceptual or identification process.

Results

The results of the experiment, presented in Figure 1, show that for each *S*, recognition improved from near chance to almost perfect performance with increases in the silent intervowel interval. The results indicate that an auditory image remained after the first vowel and was processed during the intervowel interval for recognition. The second vowel was effective in terminating the perceptual processing of the image. Figure 1 shows that the processing time necessary for asymptotic recognition differed for the 3 *Ss*. Two *Ss* (KB and MF) required about 180 msec. for optimal recognition. In contrast, the third *S* (DC) was able to identify the test vowel with only 80 msec. of additional processing time after its presentation.

These results do not necessarily estimate the duration of the preperceptual image or the perceptual processing time in typical vowel identification. Since performance reached asymptote at essentially perfect recognition, the recognition process could have been complete before the image decayed. To estimate the duration of the image, recognition accuracy must reach

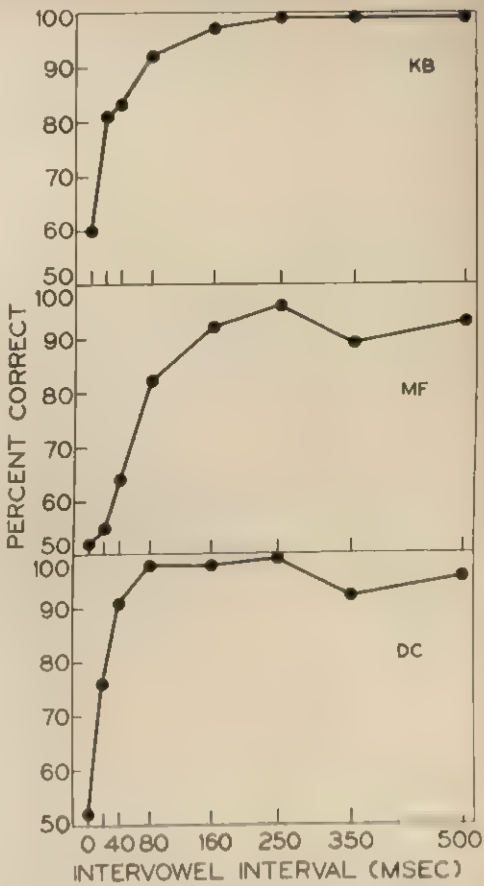


FIGURE 1. Percentage of correct test-vowel identifications for *Ss* KB, MF, and DC as a function of the duration of the silent intervowel interval: Experiment I.

asymptote below perfect performance. In the present study, the test vowel was 1 of only 2 possible vowels. With only 2 alternatives, *S* can probably make his decision very rapidly, since fewer acoustic features of the vowels need be processed for accurate recognition. With a larger number of alternatives, however, *S* would have to process more features before reaching a decision. Accordingly, increasing the number of alternatives in this task should increase the effective perceptual-processing time up to the duration of the auditory image. Accordingly, Experiment II was carried out with 4 test alternatives in the recognition masking task.

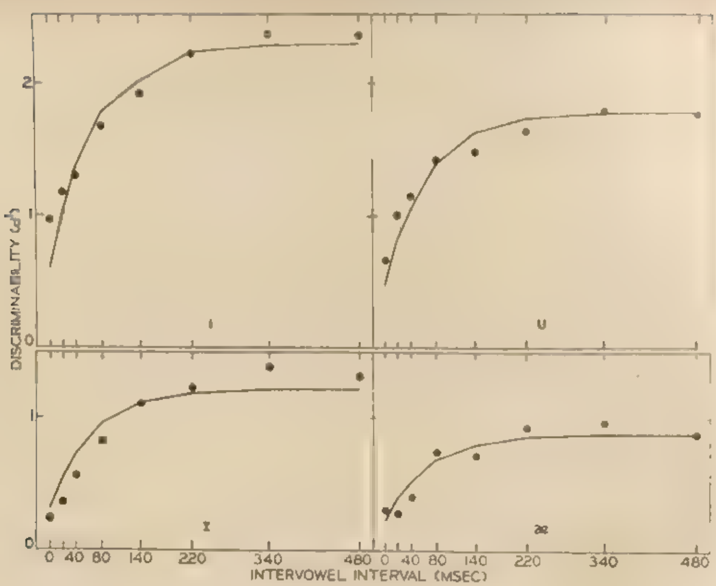


FIGURE 2. Discriminability values for the 4 test vowels as a function of the silent intervowel interval: Experiment II. (The lines are drawn through the predicted points given by Equation 1.)

EXPERIMENT II

Method

Subjects. Twelve University of Wisconsin undergraduates fulfilling a course requirement were tested for 5 days.

Procedure. In Experiment II, 4 vowels recorded by a female speaker were employed as test items. The fundamental frequency of the vowels was 192 Hz. The test alternatives were the vowels /i/ as in *heat*, /I/ as in *hit*, /a/ as in *hat*, and /U/ as in *put*. The duration of the test vowel was 20 msec. The silent intervowel interval lasted 0, 20, 40, 80, 140, 220, 340, or 480 msec. The masking stimulus was also chosen randomly from the 4 vowels, /i/, /I/, /a/, and /U/, and also lasted 20 msec.

Four *S*s were tested simultaneously in separate sound-attenuated rooms. All experimental events were controlled by a PDP-8L computer. The vowels were presented binaurally over matched headphones (Grason-Stadler TDH-49) at between 58- and 67-db. SPL. Each *S* recorded his decision by pressing 1 of 4 push buttons, labeled HEAT 1, HIT 2, HAT 3, and PUT 4, respectively. Following the 2.5-sec. response period, feedback was given by presenting the digits 1, 2, 3, or 4 on a visual display (IEE readout) for 500 msec. The intertrial interval was 2 sec. There were 2 300-trial sessions on each of 5 days. On the first day, *S* identified the test vowel without the masking vowel present. On Days 2, 3, 4, and 5 the experiment proper was carried out. Presentation of the 4 test alternatives, the 4 masking alternatives, and the 8 intervowel intervals was completely random, and each condition was programmed to occur equally often. Day 2 wa[illegible]ted as practice, so that the results are taken [illegible] the last 3 experimental days.

Results

The results were analyzed in terms of a signal detectability framework to provide a measure of vowel discriminability, d', which would be relatively independent of decision biases. It is possible to derive a measure of recognition performance for each of the 4 test vowels. The probability of identifying a vowel correctly is designated a hit, and responding with that vowel alternative to any other test alternative is designated a false alarm. For example, the probability of responding i given the test vowel /i/, $p(\mathrm{i}|\mathrm{i})$, would be a hit, whereas the probability of responding i given any other alternative, $p(\mathrm{i}|\bar{\mathrm{i}})$, would be a false alarm. The d' measure derived from these probabilities would be the index of discriminability for the vowel /i/. The d' values were obtained from the hit and false alarm rates (averaged across *S*s) in Elliot's (1964) tables.

Figure 2 presents the d' values for each test vowel as a function of the intervowel

interval. Identification of each vowel improved with increases in the silent interval between the test and masking vowels. These results can be used to estimate the duration of the preperceptual image and perceptual-processing time, since recognition reached asymptote below perfect performance. Since performance did not improve beyond an intervowel interval of 220 msec., the image could have decayed in this period. Figure 2 also shows that the test vowels were not recognized equally well, implying that the alternatives differed in discriminability. The rank ordering from least to most discriminable was /a/, /I/, /U/, and /i/.

The continuous lines in Figure 2 are drawn through the predicted points given by a perceptual-processing model (Massaro, 1970a), which assumes that readout of the information in the preperceptual image increases in a negatively accelerating manner with increases in processing time. More specifically, the perceptual strength of an item, as measured by its discriminability, d', follows an exponential growth function of time:

$$d' = \alpha(1 - e^{-\theta t}). \qquad [1]$$

The measure d' is the discriminability of the item after a presentation time of t sec. In the model, presentation time includes both the duration of the test item and the silent interval before the onset of the masking stimulus. Equation 1 indicates that the perception of a test item approaches an asymptote α at a rate of θ.

Each test vowel is assumed to function as a perceptual unit and therefore has a number of distinctive features that correspond to a feature list in long-term memory. The rate at which these features are processed is reflected in the value of θ. A vowel with a large number of distinctive features would be expected to have a large α value. A noisy or unclear vowel would have few distinctive features and a small α value. Thus, the α value can be thought of as an index of the number of discriminable features of a vowel.

According to the model, the overall differences found in identifying the test vowels should reflect differences in α, the measure of discriminability. However, θ, the rate of processing the information in the vowel, should be independent of α. In all cases, the number of features processed in any unit of time should be a constant proportion of the number of features that remain unprocessed. Equation 1 was applied to the observed results in Figure 2 by estimating a different parameter value of α for each test vowel and a single value of θ. As can be seen in Figure 2, the model describes the results fairly accurately. Thirty data points are predicted with only 5 parameter estimates. The parameter estimates of the α values for /i/, /I/, /a/, and /U/ were 2.29, 1.31, .88, and 1.77, respectively. The estimated value of θ was 15.17.

Table 1 presents the average recognition performance for the different test stimuli as a function of the masking vowel. As can be seen in the table, all masking vowels were equally effective in terminating perceptual processing of the test vowel. Of special interest is the finding that a vowel can mask itself. This means that *S* did not discriminate trials on which the test and mask were the same vowel. Since identification of the test vowel requires perceptual-processing time, it cannot be accurately compared to the masking vowel until it is recognized. The masking vowel overwrites the test vowel effectively, leaving very little information about the test-vowel presentation.

The results show that perception of a short vowel presentation improves with

TABLE 1

AVERAGE PERFORMANCE (MEASURED IN d' VALUES) FOR TEST VOWELS AS A FUNCTION OF MASKING VOWELS

Test stimulus	Masking stimulus			
	/i/	/I/	/a/	/U/
/i/	1.76	1.78	1.78	1.62
/I/	.78	.95	.96	.78
/a/	.56	.66	.23	.66
/U/	1.15	1.33	1.22	1.75
Average	1.06	1.18	1.05	1.20

increases in the silent interval after its presentation. This result provides evidence for the assumption that the duration of vowels in normal speech provides time for perceptual processing. However, increasing the duration of a vowel could also increase the amount of information in the vowel presentation. Increasing the duration of a vowel increases the information in the stimulus if one analyzes the frequency–amplitude spectrum of the steady-state vowel. With a short vowel stimulus, the distributions of energy around the respective formants are fairly large, and increasing the duration of the vowel would decrease the variance of the distributions. Accordingly, to measure the relative contribution of processing time and stimulus information in vowel perception, perception of a short vowel followed by a silent interval was compared to perception of that same vowel left on for the processing interval before presentation of the masking vowel.

Experiment III

Method

Subjects. Eight undergraduates fulfilling a course requirement at the University of Wisconsin served as *S*s for 5 days.

Procedure. The 4 test alternatives used in Experiment II were also employed in this study. The vowels were first modified digitally to give exactly the same steady-state loudness (65-db. SPL). Since the vowels had a fundamental frequency of 192 Hz., 5 fundamental segments gave a stimulus duration of 26 msec. In the continuous-processing conditions, vowels had to be presented at durations that exceeded our computer storage capacity. Accordingly, we stored 26 msec. of each vowel and repeated this segment until the desired duration was presented. The masking vowel was a 208-msec. continuous nonsense vowel made up of 2 repetitions of the 4 test vowels, /i/, /I/, /a/, and /U/, played for 26 msec. each.

On the first day of the experiment, *S*s listened to the test vowels played at durations of 208, 104, and 52 msec., respectively, in 3 successive 300-trial sessions. Feedback was given simultaneously with the test-vowel presentation. The experiment proper was carried out on the following 4 days. In the silent-processing condition, the test vowel was presented for 26 msec., followed in turn by a silent interval (0, 26, 52, 78, 130, 182, 260, or 390 msec.) and the masking vowel. In the continuous-processing condition, the test vowel was presented for 26, 52, 78, 104, 156, 208, 286, or 416 msec., followed immediately by the masking vowel. All experimental conditions (4 Test Vowels × 2 Processing Conditions × 8 Processing Intervals) were completely random in a given session and were programmed to occur equally often. Therefore, on every trial, *S* heard a test vowel of variable duration, followed in turn by a variable silent interval and a masking vowel. Following a 2-sec. response period, feedback was presented. The intertrial interval was 1 sec. All other procedural details were similar to Experiment II. Two 300-trial sessions were given, and the first 5 trials on each day were [illegible]d in the data analysis.

Results

The results shown in Figures 3, 4, 5, and 6 present discriminability values for each vowel as a function of processing time and the silent and continuous conditions. For every vowel, identification was better when the vowel was left on for the processing interval than when a 26-msec. vowel was followed by a silent-processing interval. More specifically, for all test vowels, performance reaches asymptote at a higher level of performance for the continuous- than for the silent-processing condition, suggesting that there was more stimulus information in the test vowel in the former than in the latter processing condition. Furthermore, the results also indicate that performance reaches asymptote at the same processing interval for the silent- and continuous-processing conditions. Accordingly, it appears that the rate of processing the information in the test vowel did not differ under the 2 processing conditions.

This interpretation of the results can be described in the framework of the perceptual-processing model. We first assume that a 52-msec. vowel produces more information than a 26-msec. vowel, but that the information in a vowel presentation does not increase with increases in duration beyond 52 msec. It follows that the continuous condition should only have one value of α at test vowel durations of 52 msec. or more. We assume that this α value is some constant K, times the α value for the corresponding 26-msec. vowel presentation. Accordingly, to obtain the theoretical predictions, it is necessary to estimate a different α for each of the 4 test vowels, the constant K, and a value

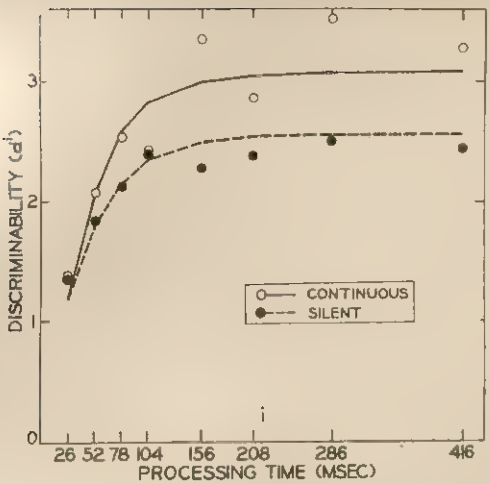


FIGURE 3. Discriminability values for the test vowel /i/ as a function of processing time during the continuous vowel presentation or the silent interval after a 26-msec. vowel presentation: Experiment III. (The lines are drawn through the predicted points given by Equation 1.)

for θ, the rate of perceptual processing, which should be constant under both the silent and continuous tone conditions.

The predicted curves are given by the continuous lines in Figures 3, 4, 5, and 6. The model does a reasonably good job of describing the data, considering the fact that 64 points are predicted by estimating only 6 parameters. The parameter values for α were 2.57, 1.41, 1.24, and 1.91 for the vowels /i/, /I/, /a/, and /U/, respectively. The estimated value for K was 1.20, and the value for θ was 23.77.

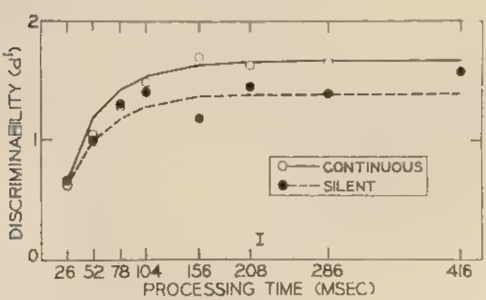


FIGURE 4. Discriminability values for the test vowel /I/ as a function of processing time during the continuous vowel presentation or the silent interval after a 26-msec. vowel presentation: Experiment III. (The lines are drawn through the predicted points given by Equation 1.)

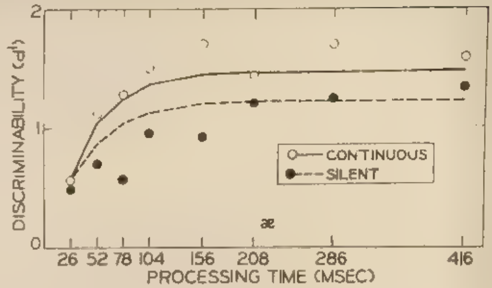


FIGURE 5. Discriminability values for the test vowel /a/ as a function of processing time during the continuous vowel presentation or the silent interval after a 26-msec. vowel presentation: Experiment III. (The lines are drawn through the predicted points given by Equation 1.)

In the silent-processing condition, recognition performance improved with increases in the silent interval, leveling off after roughly 200 msec. It might therefore be concluded that the preperceptual image of the test vowel decayed within this period. However, recognition performance reached asymptote at the same time in the continuous vowel condition. In this case, the image should not have decayed, since the test vowel was still being presented. It appears that recognition leveled off because S processed all the information available in the stimulus in the first 200 msec. In terms of the perceptual-processing model presented here, the rate of processing was

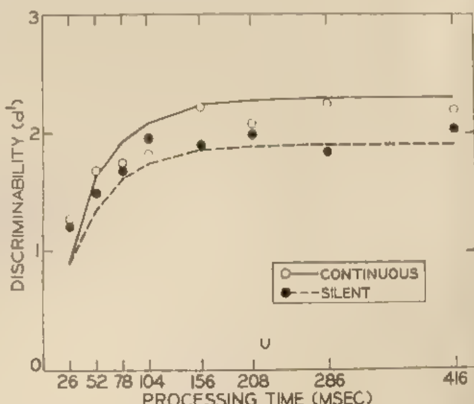


FIGURE 6. Discriminability values for the test vowel /U/ as a function of processing time during the continuous vowel presentation or the silent interval after a 26-msec. vowel presentation: Experiment III. (The lines are drawn through the predicted points given by Equation 1.)

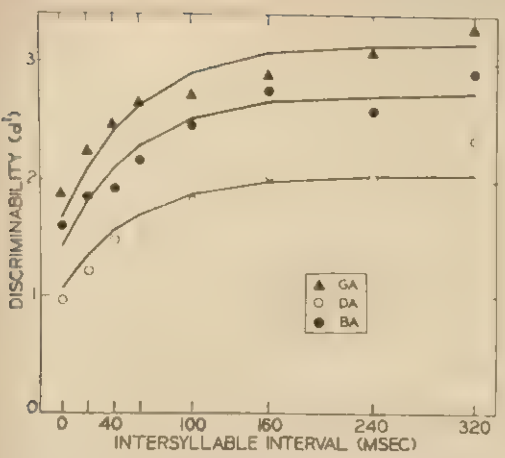


FIGURE 7. Discriminability values for the 3 test syllables as a function of the silent intersyllable interval: Experiment IV. (The lines are drawn through the predicted points given by Equation 1.)

large enough that performance reached asympote within 200 msec. It is possible that further increases in the number of test alternatives will differentially enhance performance in the continuous- relative to the silent-processing condition.

The results indicate that perceptual processing of a continuous vowel presentation continues for about 200 msec. This processing is terminated by a masking-vowel presentation. These results support the idea that the extended duration of the vowel in continuous speech allows time for perceptual processing, so that the vowel can be identified. The processing time during the steady-state vowels could also provide time for recognition of some consonants. The stop consonants are characterized by rapid transitions in the formants of the acoustic signal (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The stop consonants /g/, /d/, and /b/ can be distinguished from each other by the second formant transition, which varies as a function of place of articulation (Delattre, Liberman, & Cooper, 1955). It has been argued that the stop consonant could not be recognized while it is presented, but would require some silent time or steady-state period for recognition to take place (Massaro, 1972b).

This analysis implies that the stop consonant–vowel (CV) transition would function as a perceptual unit in speech perception. If the vowel is integrated with the CV transition, the extended duration of the vowel in normal speech would provide time for perceptual processing. Accordingly, the first part of the CV pattern probably could be identified if it is presented alone and followed by a silent interval. However, if the short CV syllable were followed by a speech sound that could not be integrated with it, perception should be disrupted, and backward recognition masking should occur. To test this, we took the CV transitions of 3 stop CV syllables and used these as test items in our backward-masking experiment.

EXPERIMENT IV

Method

Subjects. Six University of Wisconsin undergraduates fulfilling a course requirement were tested for 5 days.

Procedure. Three CVs were employed as both test and masking stimuli. The CVs were BA, DA, and GA (A as in *box*). The 2 formant CVs were synthesized at Haskins Laboratory[3] and recorded digitally for playback during the experiment. Enough of the CV syllable was presented for it to sound speechlike. Each CV syllable was of 42-msec. duration (30 msec. of transition plus 12 msec. of steady-state vowel) and was presented at 76-db. SPL. On each trial, 1 of the 3 CV syllables was presented, followed in turn by a variable interval and 1 of the 3 syllables. The silent intersyllable interval lasted 0, 20, 40, 60, 100, 160, 240, or 320 msec. Each *S* recorded his decision by pressing 1 of 3 push buttons labeled GA 1, DA 2, and BA 3, respectively. Following the 1.5-sec. response interval, feedback was provided by visually presenting the digits 1, 2, or 3 for 500 msec. The intertrial interval was 1.5 sec. Presentation of the 3 test alternatives, 3 masking alternatives, and 8 intersyllable intervals was completely random. The alternative GA could occur 2 out of 8 times and BA and DA each occurred 3 out of 8 times. All other procedural details were the same as in Experiment II.

On the first day, *Ss* identified the test syllable without the masking syllable present for 300 trials. The masking stimulus was presented in the second session. On Days 2, 3, 4, and 5, the experiment proper was carried out. There were 2 300-trial sessions per day. Only the results of Days 3, 4, and 5 were included in the data analysis.

[3] David B. Pisoni kindly provided a recording of the original unshortened syllables.

Results

The results, presented as in Figure 7, show that identification performance improved with increases in the silent intersyllable interval for all 3 CVs. The continuous lines are drawn through the predicted points given by Equation 1. Four parameter values were estimated: an α value for each CV test alternative, and a value of θ. The parameter estimates were 3.18, 2.06, and 2.75 for the values for GA, DA, and BA, respectively. The parameter estimate for the rate of perceptual processing, θ, was 17.28. Table 2 shows that the similarity between the test and masking stimuli did not systematically affect identification performance. However, BA produced less overall interference than GA or DA as a masking stimulus. This result might reflect the fact that the low frequencies of the syllable BA are less effective in terminating perceptual processing of an earlier CV presentation.

TABLE 2

AVERAGE RECOGNITION PERFORMANCE (MEASURED IN d' VALUES) FOR TEST SYLLABLES AS A FUNCTION OF MASKING SYLLABLES

Test syllable	Masking syllable		
	GA	DA	BA
GA	3.13	2.48	2.32
DA	1.40	1.58	1.94
BA	1.66	2.38	3.18
Average	2.10	2.22	2.72

DISCUSSION

These studies were carried out to determine the size of perceptual units employed in speech perception. It was argued that the perceptual unit could not be larger than the temporal life of a preperceptual image of the speech stimulus. If the information in the image decays before it is synthesized, recognition would not be possible. The results indicate that processing the speech stimuli employed here occurs within 200–250 msec. after the stimulus presentation. Accordingly, some transformation of the speech signal must occur about every ¼ sec. The stimulus within this time period must function as a perceptual unit. There must be sufficient information in this signal (and in the context in normal speech) for contact with information in long-term memory. To the extent the present studies simulate the processing of real speech, it might be concluded that the first stage of speech recognition must operate on acoustic segments that are of less than 250-msec. duration. If this is the case, perceptual units appear to be of roughly syllabic length, as suggested by Huggins (1964) and Massaro (1972b).

There are at least 2 apparent difficulties with assuming that the first stage of recognition involves a transformation or synthesis within 250 msec. after a sound pattern is presented. The first is that the utilization of auditory information occurs across much longer temporal periods. The memory for the quality of a speaker's voice requires an auditory memory that can last indefinitely. The interpretation of prosodic information such as stress requires an auditory memory that lasts for the length of a phrase or sentence. Most importantly, the recognition of sound patterns depends on an auditory memory for those sound patterns or even other sound patterns by the same speaker. For example, Ladefoged and Broadbent (1957) showed that the acoustic characteristics of some vowels of a speaker influenced the recognition of other vowels by the same speaker. Accordingly, the listener must have remembered some of the auditory characteristics of the vowels for longer than 250 msec. The auditory memory illustrated by these examples, however, is not preperceptual, but is auditory information that has been actively synthesized.

Massaro (1972a, 1972b) has distinguished between two stages of auditory information processing which involve preperceptual and synthesized auditory storage, respectively. The first stage involves recognition of a sound pattern held in a preperceptual auditory image. The recognition of the sound pattern produces a synthesized auditory percept in synthesized auditory storage. Synthesized auditory storage has a much longer life span than preperceptual images and could be responsible for auditory memory effects that operate over longer time periods than are given by preperceptual auditory storage. Accordingly, the auditory storage responsible for memory for voice quality, the Ladefoged and Broadbent (1957) effect, and for interpreting prosodic features of a sentence, is not preperceptual, but is at the level of synthesized auditory storage. (For a more complete discussion of the differences in pre-

perceptual and synthesized auditory storage, see Massaro, 1972a, 1972b.)

The second problem is that it is well known that the acoustic sound pattern corresponding to a syllable changes with changes in the speaker, the rate of speaking, the stress patterns, and (most importantly) the influence of neighboring syllables (Fant, 1962; Ohman, 1966). However, this problem is not unique to speech recognition, but is also the major deterrent to the development of a recognition device for handwritten text. This observation tells us that the representation of a perceptual unit in long-term memory must be a highly flexible one, with a number of normalization algorithms that operate on the sound-pattern input. Luckily, in normal speech, context or redundancy significantly reduces the number of valid alternatives for each sound pattern. The sound pattern can, therefore, be noisy, since redundancy reduces the number of acoustic features that are necessary to recognize the pattern correctly.

REFERENCES

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transition cues for consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 769–773.

Elliot, P. B. Tables of d′. In J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964.

Fant, C. G. M. Descriptive analysis of the acoustic aspects of speech. *Logos*, 1962, **5**, 3–17.

Fletcher, H. *Speech and hearing in communications.* Princeton, N.J.: D. Van Nostrand, 1953.

Gray, G. W. Phonemic microtomy: The minimum duration of perceptible speech sounds. *Speech Monographs*, 1942, **9**, 75–90.

House, A. A. On vowel duration in English. *Journal of the Acoustical Society of A*[illegible]*a*, 1961, **33**, 1174–1178.

Huggins, A. W. F. Distortion of [illegible]mporal pattern of speech: Interruption and [illegible]rnation. *Journal of the Acoustical Society of A*[illegible]*a*, 1964, **36**, 1055–1064.

Ladefoged, P., & Broadbent, D. E. [illegible]mation conveyed by vowels. *Journal of* [illegible]*oustical Society of America*, 1957, **29**, 98–10[illegible]

Liberman, A. M., Cooper, F. A., Sh[illegible]eiler, D. P., & Studdert-Kennedy, M. [illegible]eption of the speech code. *Psychological* [illegible] 1967, **74**, 431–461.

Massaro, D. W. Perceptual process[illegible]d forgetting in memory tasks. *Psycholog*[illegible] *Review*, 1970, **77**, 557–567. (a)

Massaro, D. W. Preperceptual audit[illegible]nages. *Journal of Experimental Psycholog*[illegible], [illegible], **85**, 411–417. (b)

Massaro, D. W. Effect of masking ton[illegible]ration on preperceptual auditory images. *J*[illegible]*al of Experimental Psychology*, 1971, **87**, 14[illegible]

Massaro, D. W. *Preperceptual and synthes*[illegible] *auditory storage.* (Wisconsin Studies in H[illegible]n Information Processing Rep. 72-1) Madis[illegible] University of Wisconsin, Wisconsin Math[illegible]atical Psychology Program, November 1972.

Massaro, D. W. Preperceptual images, pr[illegible]ssing time, and perceptual units in auditory perception. *Psychological Review*, 1972, **79**, 124–145. (b)

Massaro, D. W. Stimulus information versus processing time in auditory pattern recognition. *Perception & Psychophysics*, 1972, **12**, 50–56. (c)

Ohman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 1966, **39**, 151–168.

Peterson, G. E., & Lehiste, I. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 1960, **32**, 693–703.

Suen, C. Y., & Beddoes, M. P. Discrimination of vowel sounds of very short duration. *Perception & Psychophysics*, 1972, **11**, 417–419.

(Received January 13, 1973)

ournal of Experimental Psychology
1974, Vol. 102, No. 2, 209–214

EFFECTS OF SAME–DIFFERENT PATTERNS ON TACHISTOSCOPIC RECOGNITION OF LETTERS[1]

ROBERT P. INGALLS[2]

California State University, Chico

Two experiments investigated whether *S*s recognize that letters are physically identical prior to naming them, or whether the processes of naming and physical matching occur in parallel. Stimuli consisting of 3 letters were presented tachistoscopically at a duration such that *S*s could identify all 3 letters about half the time. The *S*'s task was to name the letters. Some stimuli had all 3 letters the same, some had all 3 letters different, and some had 2 the same and 1 different. There was no difference in recognition performance between the condition with all 3 letters the same and the condition with all 3 letters different, suggesting that the physical matching process and the naming process are not sequential. However, recognition of letter patterns with the first 2 letters the same and the last letter different was much lower. The results were replicated with number stimuli.

A given pattern can be categorized in numerous different ways. For example, the pattern A can be processed as a letter with a certain sound, as a vowel, as a part of a word, or simply as a physical pattern consisting of 2 oblique lines with a horizontal line connecting them. A series of experiments by Posner and Mitchell (1967) showed that it takes longer to perform some categorizations than it does to perform others. Posner and Mitchell asked *S*s to decide whether 2 letters were the same or different, with the dependent variable being reaction time, and they found that physical matches took less time than did name matches, and name matches took less time than did matches based on the consonant–vowel distinction. A question which immediately comes to mind is how the various levels of categorization are related. There are 2 basic possibilities: they can be connected serially (sequentially), that is, physical processing is done first, and the information from physical processing is used to perform higher level categorizations such as naming; or they can be connected in parallel, essentially independently of one another.

The position that the 2 levels are sequential was taken in a study by Beller (1970), who based his thinking on Neisser's (1967) theory. Beller equates Posner and Mitchell's (1967) Level 1 processing to Neisser's concept of preattentive mechanisms, while Level 2 corresponds to figural synthesis and pattern recognition. To verify his approach, Beller used the assumption that preattentive processes are carried out in parallel. If Level 1 (physical match) is done in parallel, then it should not matter how many items were presented; the time needed to classify them all as same or different based on a physical match would be the same. His stimuli consisted of 2, 4, or 8 letters which *S* had to classify as same or different according to their name, as in Posner and Mitchell's Level 2 instructions. There were 2 interesting conclusions to his study. The first was that the times required to make physical matches were the same regardless of the number of letters in the stimulus; i.e., it took no longer to classify 8 letters as physically identical than it did to classify 2. These data provide strong support for the assumption of spatiotemporal parallelism in Node 1 processing. The second conclusion was that when all the stimuli were physically identical, *same* response times were faster than

[1] This research is based on a doctoral dissertation submitted to the University of Connecticut. The author is indebted to his committee, Michael Turvey, David Zeaman, and Donald Dickerson, for their invaluable advice and assistance throughout. The equipment was purchased through University of Connecticut Research Foundation Grant 102.

[2] Requests for reprints should be sent to Robert P. Ingalls, who is now at Russell Sage College, Troy, New York 12180.

same response times when the stimuli differed physically but had a common name, even though the instructions in all cases were to judge on the basis of name similarity rather than physical similarity. This observation supports Beller's contention that the 2 levels are operationally serial because, apparently, once *S* recognizes that the letters are physically identical, he does not have to name each one even though his instructional set is to do so.

The purpose of the present experiments is to investigate the question of whether the physical match subsystem necessarily precedes the naming process. The general procedure was to present tachistoscopically a series of stimuli, each consisting of a 3-letter (or 3-number) array, followed by a masking stimulus. The time between the onset of the stimulus pattern and the onset of the mask was manipulated so that each *S* was able to report the 3 letters correctly about half the time. Previous research indicates that the amount of time that *S* has available to read out information from the stimulus is determined by the interval between the onset of the stimulus and the onset of the mask (see Kahneman, 1968, for a general review). Of course, this does not mean that no processing occurs after the mask comes on but rather that raw stimulus information is no longer available. Some of the stimuli had all 3 items the same, others had all 3 items different, and others had 2 the same and 1 different. The dependent variable was the percentage of letters correctly identified in each condition. If physical matching is an integral part of the preattentive processes as suggested by Beller (1970) and Neisser (1967), performance in the condition with all 3 letters the same should be better than performance in the case with all 3 letters different, with the other conditions in between these 2, because the preattentive mechanism would first detect the physical similarity of the 3 letters. This information would facilitate recognition because *S* would only have to identify any 1 of the 3 letters in order to get all 3 correct. If, however, the physical matching level and the naming process operate in parallel, then there should be no difference between the various conditions.

Experiment I

Method

Design. The *S*'s task was to identify [illegible] series of stimuli, each consisting of 3 letters, pre[illegible]ted on a Gerbrands 3-channel tachistoscope. Th[illegible] were 5 conditions in the experiment. In [illegible]ndition (AAA) all 3 letters were the same [illegible] econd condition (AAB) the first 2 letter[illegible] same and the last letter was different. I[illegible]d condition (BAA) the last 2 letters were [illegible] with the first letter different. In the fou[illegible]dition (ABA) the first and third letters were [illegible]e and the middle letter was different. In the [illegible]dition (ABC) all 3 letters were different. Ea[illegible]eived all stimuli in all conditions. The depend[illegible]riable was the proportion of letters correctly [illegible]ned in each of the 5 conditions.

Subjects. The *S*s were 15 undergrad[illegible]t the University of Connecticut who participated in the experiment for approximately ½ hr. each for course credit.

Apparatus and procedure. The apparatus was a Gerbrands T-3B-1 3-channel tachistoscope. Each trial was initiated by *S* pressing the start button. Immediately, the first field came on for 500 msec.; it consisted of 2 horizontal lines centered in the display, which indicated where the letters were going to be. The luminance was .1 ftL. as measured by a Spectra brightness spot meter. The second field came on immediately at the offset of the first and contained the letters to be identified. The duration for the second field varied between *S*s and [illegible] determined in the pretraining as the duration at which *S* was making 50% correct responses. The mean display time for the second field was 68 msec. with a standard deviation of 13.6. The third field was a mask made up of a series of interlocking circles and was in the same location as the letters. It came on immediately at the offset of the second field and stayed on for 50 msec. The luminances of the second and third fields were identical at 1.5 ftL.

Stimuli. The stimuli consisted of 50 4 × 6 in. cards, each of which contained 3 capital letters produced by Letraset 48 pt. Futura Medium. Only 10 of the 26 letters were used, C, D, H, L, M, N, P, R, S, and T. In the apparatus, each letter filled approximately 46′ of visual angle, and a display of 3 letters had a visual angle of approximately 2°51′. Each letter was used equally often in each condition to the extent that this was possible. There were 10 stimuli in Condition AAA, 5 in Condition AAB, 5 in Condition BAA, 10 in Condition ABA, and 20 in Condition ABC. The stimuli were divided randomly into blocks of 5 each, each block containing 1 of each condition to the extent that this was possible. The 10 blocks were then ordered randomly and each *S* was begun on a different block. In addition each *S* was given pretraining on stimuli identical to those used in the experiment, with the same proportion of each condition as in the actual experiment; i.e., 40% of the pretraining stimuli were of the ABC type, 20% were AAA pattern, 10% were BAA, etc.

[illegible]he purpose of the pretraining was to stabilize the [illegible]ance and to determine for each S that dura[illegible] which his performance was 50%. The num[illegible] of pretraining trials needed for this varied, with [illegible]an of 48.9 pretraining trials per S (SD = 12.4). [illegible] S said his response aloud immediately at the [illegible] each trial. The Ss were required to guess if [illegible] were not sure of the answer, and they had been [illegible]old before the experiment which 10 letters were to be used. If S's response was wrong, he was told what the correct letters had been. The purpose of this feedback was to acquaint him with the possible combinations of letters and the approximate proportions of each. It had been determined in pilot work that Ss tended to guess 3 different letters each time without this feedback, and so it was necessary to indicate that some stimuli might have all 3 letters the same and that others might have 2 the same.

Results

Figure 1 shows the mean percentage correct for each condition and position of the letter (first, second, or last). Averaging across positions, although performance in the AAA condition was slightly better than performance in the ABC condition as predicted by the Beller (1970) – Neisser (1967) model, this difference was negligible. The only stimulus condition which appeared to differ widely from the others was the AAB condition, in which performance was inferior. An analysis of variance indicated that both main effects of position and condition and the interaction between them were significant: for condition, F (4, 196) = 12.41, $p < .001$, for position, F (2, 196) = 51.46, $p < .001$, and for the interaction, F (8, 196) = 2.93, $p < .01$.

A Tukey test was done to determine which conditions differed significantly from each other. The only finding here was that the AAB condition differed reliably from the other 4 conditions at the .01 level. The difference between the AAA and ABC conditions did not even approach significance. Another Tukey test on paired comparisons between positions averaged across condition indicated that performance in the first position differed from performance in the other 2 at the .01 level, while the difference between Position 2 and Position 3 approached significance at the .05 level. The significant interaction appears to be due primarily to the fact that the condition effect (performance in the AAB condition being inferior to performance in the other conditions) was found only in the last 2 positions.

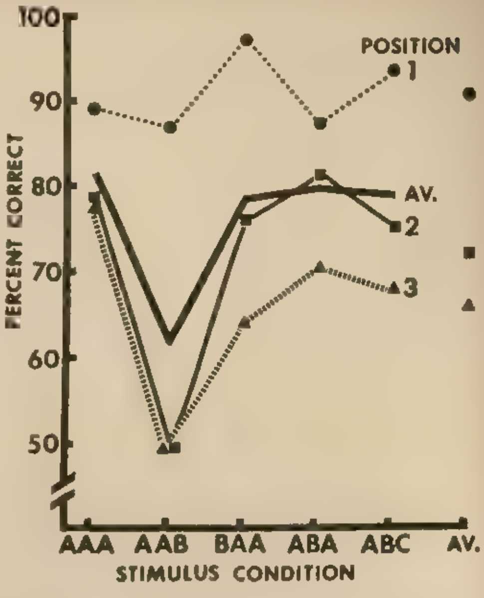


Figure 1. Percentage of letters correctly identified for each condition and position, Experiment I. (The A–B combinations represent the 5 same–different patterns. Av. = average.)

An analysis of the kinds of errors that were made was done to see if it would provide further clues regarding the mental processes involved in this task. The results are shown in Table 1. If the Beller (1970) model is taken literally, Ss should be able to detect the same–different pattern of the stimulus (i.e., all letters the same; first 2 letters same, last different; or whatever) before identifying the actual letters. Thus, there should be many cases where the response had the correct same–different pattern but missed one or more of the letters. The error analysis indicated that this was not the case. Many errors involved Ss getting all the letters correct but in the wrong order (17% in the ABC condition, 7% in the ABA condition, 20% in the AAB condition, and 2% in the BAA condition); while there were fewer cases in which the same–different pattern was correct but one or more of the letters was missed (13% in the ABA condition, 11% in the AAB condition, and none in the BAA condition). In the AAA condition there was not a single

TABLE 1

Frequency of Types of Errors in the Five Conditions of Experiment I

AAA		AAB		BAA		ABA		ABC	
Type of error	No.	Type of error	No.	Type of error	No.	Type of error	No.	Type of error	No.
XXX	0	AAA	6	BBB	1	AAA	3	[illegible]	10
AAX	12	AAX	6	BBA	5	AXA	7	[illegible]	6
XAA	3	ABB	1	BAB	10	BXB	1	[illegible]	2
AXX	2	ABA	11	ABA	1	AAB	4	[illegible]	2
AXA	18	BAB	1	BXB	1	BBA	1	[illegible]	2
XAX	2	ABX	9	BAX	13	ABX	23	[illegible]	1
AXY	9	AXB	10	BXA	9	XAB	6	[illegible]	4
XAY	8	XAB	5	Other	2	XBA	5	[illegible]	3
Other	3	Other	6			Other	10	[illegible]	1
					42 (total)			[illegible]	4
	57 (total)		55 (total)				60 (total)	BBC	1
								ABX	38
								ACB	19
								AXC	13
								BAC	3
								Other	19
									129 (total)

Note. The A-B combinations heading the columns represent 5 same-different patterns determining the 5 conditions.

case where S knew that all 3 letters were the same, but did not know what the letter was (an XXX error).

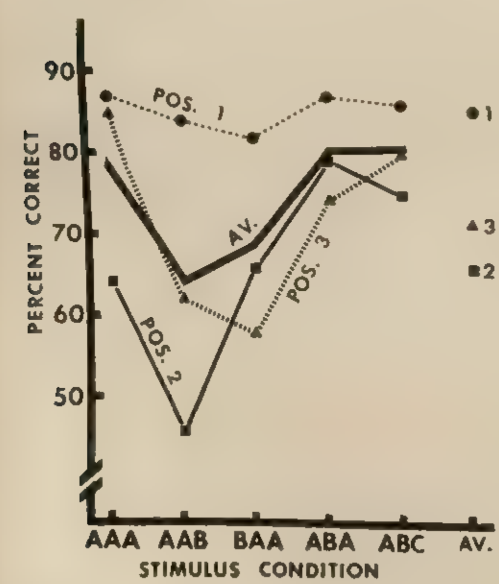


Figure 2. Percentage of letters correctly identified for each condition and position, Experiment II. (The A-B combinations represent the 5 same-different patterns. Av. = average.)

Experiment II

There were 2 major findings in Experiment I that could be viewed with some suspicion. First, the absence of a significant difference between the AAA and ABC conditions supports the null hypothesis, which is a questionable conclusion. Second, the poor performance of the AAB condition was so unexpected that it warranted replication.

Method

The design, equipment, and procedure of the second experiment were identical to those of Experiment I with the following differences: the stimuli were numbers instead of letters; there were a total of 40 stimuli instead of 50; Conditions AAA, ABA, and ABC each had 10 stimuli and Conditions AAB and BAA each had 5 stimuli; and only 10 *Ss* were used instead of 15.

Results

The results of Experiment II are shown in Figure 2. The 2 tenuous conclusions of Experiment I were firmly supported by these data. First, the performance of the AAB condition was inferior to the performances in the other conditions; and second,

there was little difference between the performances of the AAA and ABC conditions, this case the ABC condition was ior to the AAA condition. This difference is in the opposite direction from that licted from the Beller (1970) – Neisser 57) model. An analysis of variance indicated that both main effects of condition and position were significant at the .001 level and the interaction between them was significant at the .05 level: for conditions, $F(4, 126) = 5.36$, for position, $F(2, 126) = 15.63$, and for the interaction, $F(8, 126) = 2.05$. A Tukey test on the condition means averaged across positions indicated that the AAB condition differed from the ABC and ABA conditions at the .01 level and from the AAA condition at the .05 level. Another Tukey test on the position means averaged across conditions indicated that the first position performance differed from the other 2 at the .01 level.

Discussion

The most important conclusion of this research is that the recognition of AAA stimuli, in which all 3 letters were the same, and the ABC stimuli in which all 3 letters were different, did not differ significantly, and this was true with either numbers or letters as stimuli. It can be concluded from this finding that either the same–different pattern of the stimuli was not discerned before the letters were identified or that it was discerned but not used. The latter seems unlikely. If the same–different pattern had been detected prior to the letters being identified, then there should have been many more cases in the AAA condition in which *S*'s response had all 3 letters the same but used the wrong letter. This result calls into question Beller's (1970) assertion that physical sameness is detected by the preattentive mechanisms prior to identification.

However, it is still necessary to resolve the difference between Beller's (1970) data and the present data. Beller found that when *S*s were asked to make same–different judgments on the basis of name, response times were considerably faster when all stimuli were physically identical, and it took no longer to classify 8 physically identical stimuli as same than it did to classify 2. He concluded that physical match was done prior to the actual naming process and that it was done in parallel. The present study concluded exactly the opposite. However, there are some important differences between the task of Beller's *S*s and the task of the *S*s in the present experiments. Beller's *S*s were required to decide whether the letters presented had the same name or not. The actual task of naming was only implicit. In the above experiment, *S*s were required to actually say the names of the letters. Apparently this difference can produce important differences in processing strategies in the sense that certain preattentive processes seem to be optional. When *S*s are required to make a name match, i.e., say whether 2 or more letters have the same name or not, they apparently employ a preattentive mechanism which makes physical matches prior to actually naming the letters. However, when their only assignment is to name the letters, then they can somehow forgo the use of the matching mechanism. In other words, which preattentive mechanisms are employed seems to depend on the specific task.

A second finding of the present experiments is that recognition performance with AAB stimuli was considerably lower than performance in other conditions. This result is puzzling; perhaps some sort of symmetry notion can account for it; symmetry has been found to play a role in other perception experiments (e.g., Bryden, 1968), but this would require some stretching of the notion of symmetry. At present, this phenomenon will have to be left unexplained.

A third finding was that *S*s were much more likely to get the first letter correct than the other 2 letters. This finding is not surprising; it has been reported in other studies (Bryden, 1960). What it indicates is that *S*s are most probably reading from left to right, and it provides support for the notion that the letters are synthesized serially, one after the other, rather than in parallel (although the observation that *S*s often knew what letters were present without knowing their order suggests the opposite). However, a parallel processing model could also account for the position effect by postulating that since the letters are spoken in a left-to-right order, there is a greater chance that *S* will forget 1 of the last 2 letters rather than the first letter because he has to keep them in memory while he reports the first.

Finally, these experiments may be taken as a demonstration of a case where one type of meaning information can be extracted directly from a stimulus without first having to

attend to, recognize, or identify a lower and presumably more basic type of physical or sensory information. Other research suggests that *S*s are capable of performing in a similar manner with other types of information. For example, Brand (1971) reported that some *S*s were able to search through a list of letters more rapidly for "any number" than for a specified letter, suggesting that they are capable of determining whether a pattern is a number or a letter without first identifying it. This is not to say that basic sensory information is of no importance to perception, but rather that there is some sort of higher order invariant of "numberness" or "letterness" which can be detected independently of identifying any particular alphanumeric character. Similar notions have been expressed by Gibson (1966). In any event Brand's finding agrees with the present results in that it suggests that what is presumed to be higher order information can be extracted without first identifying lower order information. Thus, simple hierarchical models of pattern recognition in which lower level information is extracted first and then used to detect higher order features must be called into ... question.

REFERENCES

Beller, H. K. Parallel and serial stag... ing. *Journal of Experimental Psycholo...*, 84, 213–219.

Brand, J. Classification without identi... visual search. *Quarterly Journal of Exp...* *Psychology*, 1971, 23, 178–186.

Bryden, M. P. Tachistoscopic recognition ... alphabetic material. *Canadian Journal o...ogy*, 1960, 14, 78–86.

Bryden, M. P. Symmetry of letters as a f... tachistoscopic recognition. *American Jou...* *Psychology*, 1968, 81, 513–524.

Gibson, J. J. *The senses considered as pe... systems.* Boston: Houghton Mifflin, 1966.

Kahneman, D. Method, findings, and theo... studies of visual masking. *Psychological Bu...* 1968, 70, 404–425.

Neisser, U. *Cognitive psychology.* New Y... Appleton-Century-Crofts, 1967.

Posner, M. I., & Mitchell, R. F. Chronom... analysis of classification. *Psychological Re...* 1967, 74, 392–409.

(Received February 6, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 215–225

ACQUISITION AND TRANSFER OF INTERCONTROL MOVEMENT DEPENDENCE IN TWO-DIMENSIONAL COMPENSATORY TRACKING[1]

RICHARD B. HOPPE[2]

University of Minnesota

Two experiments on the acquisition and transfer of dependent (D) and independent (I) interaxis control movement patterns are reported. Task dependence, defined as the correlation between the stimulus signal rate functions in the 2 axes, was varied in a transfer design. Twelve *Ss* in Experiment I and 24 *Ss* in Experiment II trained on either an I or a D task then were transferred to a test task with the same (Experiment I) or different (Experiment II) intraaxis characteristics, and the same (nonshift) or different (shift) interaxis dependence. Results showed the following: (*a*) *Ss* approached the dependence required by the task in training trials, (*b*) shift *Ss* had greater error increases in test than nonshift *Ss*, (*c*) *Ss* trained on an I task and tested on a D task (I–D *Ss*) showed greater relative negative transfer than D–I *Ss*, and (*d*) the fit of the multiplicative rule for time-on-target scores to observed data was not related to movement dependence or to movement simultaneity.

The principal objects of this research were (*a*) to ascertain the relationship between the 2-hand movement dependence exhibited by *Ss* on a 2-dimensional tracking task and the fit of the multiplicative rule for time-on-target (TOT) scores, and (*b*) to determine whether the interlimb movement dependence required by the task is necessary to an adequate description of what *Ss* learn in tracking. The multiplicative rule for TOT scores is an empirical relationship between the product of part-task scores and observed whole-task performance. Ellson (1947) first noted that the proportions of TOT in each of the dimensions of a multidimensional task appeared to be statistically independent, so that

$$P(\mathrm{TOT}_{abc}) = P(\mathrm{TOT}_a) \times P(\mathrm{TOT}_b) \times P(\mathrm{TOT}_c), \quad [1]$$

where $P(\mathrm{TOT}_{abc})$ is the proportion of time that *S* was on target in all 3 dimensions simultaneously, and $P(\mathrm{TOT}_a)$, $P(\mathrm{TOT}_b)$, and $P(\mathrm{TOT}_c)$ are the proportions of TOT in each of the 3 part tasks measured without regard to each other. This rule has been shown to be very powerful. In a variety of tracking tasks, both pursuit and compensatory, and for various stages of practice, the rule predicts observed simultaneous $P(\mathrm{TOT})$ within approximately 2% (Adams & Webber, 1961; Bilodeau, 1955, 1957; Bilodeau & Bilodeau, 1954; Dick, 1970). The rule is robust over target sizes (Adams & Webber, 1961), at least under certain circumstances, and for up to 4-dimensional tasks (Bilodeau, 1957). As an empirical law relating part-task scores to whole-task performance in terms of system error, the multiplicative rule is unrivaled in its precision and range.

The question at issue is whether the operation of the multiplicative rule implies that the behavior giving rise to the TOT scores is somehow independent in the

[1] The research reported in this article is drawn from a dissertation submitted to the Graduate School, University of Minnesota, and was supported in part under Contract NOOO14-68-A-0141-003 between the University of Minnesota and the U. S. Office of Naval Research. Preparation of this article was supported in part by grants to the University of Minnesota, Center for Research in Human Learning, from the National Science Foundation (GB-17590), the National Institute of Child Health and Human Development (HD-01136 and HD-00098), and the Graduate School of the University of Minnesota. The author wishes to thank P. W. Fox, who supervised the dissertation research, and Keith Hazelton, who provided computer programming support.

[2] Requests for reprints should be sent to Richard B. Hoppe, who is now at the Department of Psychology, Kenyon College, Gambier, Ohio 43022.

several control dimensions. This implication is based on the similarity of the form of the multiplicative rule to the equation defining the joint probability of independent events. Adams and Webber (1961) explicitly presumed this to be the case when they said,

> Therefore, at present, we have no evidence to contradict the independence law [multiplicative rule] and must accept it as a first quantitative principle of complex tracking *behavior* with no known restrictions on task, practice, or characteristics of the input signals [p. 200, italics added].

Other investigators have been more aware of the difficulty of inferring behavioral conclusions from TOT error data. In their discussion of an experiment in which the data were well-predicted by the multiplicative rule, Bilodeau and Bilodeau (1954) said,

> To conclude from these data that the relation [between *P*(TOT) scores] is essentially zero, for example, is not necessarily to conclude that manipulative corrections between components are always independently performed [p. 42].

Nevertheless, they went on to say that "over a sizable time interval the events [manipulative corrections] can on the whole be considered to be independent [p. 42]," this conclusion being based on *P*(TOT) data.

The principal difficulty with these kinds of inferences is the meaning to be assigned to "independence" as it refers to behavior and to response patterns. The term has taken several meanings in the literature. Perhaps the most commonly encountered meaning identifies the degree of independence (or dependence) with time-sharing patterns across control dimensions. If *S* moves 2 controls in an alternating manner, he is displaying little time-sharing; if he tends to move the controls simultaneously, he is displaying increased time-sharing. This definition of independence is apparently assumed in the Bilodeau and Bilodeau (1954) quotations above, though, as noted, they depend upon TOT error data for their premises. The term has taken other meanings. For example, Adams and Webber (1961), also quoted above, defined behavioral independence in terms of cross-correlation functions relating tracking error signals in 2 dimensions of tracking. A misconception underlying these definitions is that tracking error is a valid index of *S*'s responding on the task. As both Dick (1970) and Hoppe (1972) have pointed out, tracking error confounds the system forcing function and *S*'s responses. As a consequence, without an adequate description of the forcing function it is impossible to separate their effects. Error indices are measures of system performance, not human responding, and without a behaviorally relevant description of the task set for *S* it is not possible to meaningfully assess his responding based upon error data.

Dick's (1970) major contribution was to (correctly) observe that conclusions about human responding in a tracking system must be based upon indices that are unambiguously descriptive of that responding. Accordingly, Dick performed his major analyses on the recorded output of *S*s' control manipulanda. While Dick's insight was notable, there is one more indispensable prerequisite to an experimental methodology adequate for the study of human skilled behavior. In order to be interpretable, any measure must have a referent. The referent for measures of tracking error is zero system error. The referent for measures of tracking behavior is not so obvious. It is what Hoppe (1972) has called the reference response pattern (RRP). The RRP is the precise spatiotemporal pattern of control movements necessary to produce zero system error through the length of a trial. The RRP derives from a reconceptualization of *S*s' task as a behavior-reproduction task, at least for the purposes of measurement, and serves the same purpose as the zero error referent for the error measures. (The RRP is developed in more detail in Hoppe, 1972.) The assumption of zero error as the criterion system performance implies that *E* should use indices of deviation from zero system error as dependent variables. The assumption of the RRP as the criterion behavior patterning implies that *E* should use indices that reflect deviations from that pattern as dependent variables. The

two are not mutually exclusive, but are complementa

As a consequence, a prerequisite for tracking research is a specification of the task requirements in the same terms to be used to describe Ss' behavior. Since the present research deals with response dependence, it is necessary to define dependence in a way that renders it quantifiable and experimentally manipulable. In this research, dependence is defined as the relationship of the rate and direction of control movement in one axis to the rate and direction of movement in the other axis. More precisely, interaxis movement dependence in a 2-dimensional tracking task is indexed by the product-moment correlation of the rate functions in the 2 axes, the 2 rate functions being synchronized in time. This correlation is the dependence coefficient (DC) and is an index of what kind of movements of one control go with what kind of movements of the other control.

consider a 2-dimensional task in which S controls Dimension A by pushing lling a control rod with his right hand. l controls Dimension B by l-grip handle held in his he definition above, S is splaying dependent responding when the DC derived from his response movements departs from zero. This is equivalent to saying that S is moving in a dependent manner when, e.g., a clockwise twist of the left-hand control is statistically associated with pushing on the right-hand control. (Zero dependence does not mean that S is not moving both controls continuously and/or simultaneously, but only that there is no statistical association of the rates and directions of movement of the 2 controls.) The sign of DC reflects the particular association; in the research reported in this article, a negative DC indicated an association of counterclockwise twisting with pushing and clockwise twisting with pulling. One question addressed by the research reported here is whether task dependence so defined, when varied, produces concomitant variation in the fit of the multiplicative rule to Ss' error data.

The second and more general question addressed is whether dependence so defined is part of what Ss learn when practicing a tracking task. Textbook accounts of skills learning typically refer to the integration of part skills, or to the spatiotemporal patterning of movements (see, e.g., Fitts & Posner, 1967; Kay, 1969; O'Donnell, 1969), but there are few experimental data which give these assertions more than intuitively based meaning. In characterizing the task in the same terms to be used to describe Ss' behavior in multidimensional tracking, the present research aims to provide the beginning of an empirical substrate for an account of learning of complex psychomotor skills.

Experiments I and II

Two experiments were performed to assess the effects of varying the DC task on (a) overall system performance, indexed by root mean square (rms) error, (b) the goodness of fit of the multiplicative rule to Ss' data, and (c) the degree of interaxis control movement dependence displayed by Ss. The experiments were similar in general form and will therefore be described together. Since the effort in the experiments was to determine the importance of task dependence to a description of what S has learned, a transfer design was used; Ss were trained on one task dependence and tested on another. In this manner it was hoped to ascertain whether Ss were merely coming to display the interaxis control dependence required by the task, or were in fact learning something about that dependence that has effects on subsequent responding on a task with a different DC. During training on a dependent (D) task, the DC of Ss' responding is expected to approximate, over trials of practice, the DC defined as criterion by the system forcing functions. This would demonstrate that Ss can impose a D structure on their responding if the task requires it. Transfer to an independent (I) task is expected to show, by comparison with an appropriate control, whether the D response patterning is a learning or performance variable. It is possible that Ss in a D task are learning to

respond appropriately in the 2 axes simultaneously, but are not learning the dependence between responses in the 2 axes. If these *S*s are shifted to an I task that is identical to the D task except for the dependence between axes, *S*s' DC should shift immediately to the test task value, and the system performance measure should show no change as a function of the change in tasks. The same prediction would hold for *S*s trained on an I task and tested on a D task. If, on the other hand, there is an effect due to the shift in tasks, then one may conclude that *S*s were learning the dependence of responding in the 2 axes.

The experiments were also designed to assess the relationship between the multiplicative rule for TOT scores and behavioral independence/dependence. If the rule is in fact an index of behavioral independence as defined by DC, then it is to be expected that it will differentiate between *S*s trained on tasks having different DCs. Specifically, the rule should predict the *P*(TOT) scores of *S*s performing on an I task better than the scores of those performing on a D task. If there is no regular relationship between behavioral dependence and the rule's predictions, then one may conclude that the rule is not an indicant of independent responding on the 2 controls.

The general approach in the experiments was to train *S*s on a 2-dimensional compensatory tracking task with a known, *E*-determined DC. The *S*s were then transferred to a task with either identical (Experiment I) or different (Experiment II) intraaxis characteristics, and the same (nonshift groups) or different (shift groups) interaxis control movement dependence. The experiments were full transfer designs, with all 4 groups (I–I, D–D, I–D, and D–I) run in both.

Method

Subjects. The *S*s were 36 right-handed male undergraduate students in introductory psychology classes at the University of Minnesota. Experiment I used 12 *S*s, and Experiment II used 24 *S*s; all *S*s were volunteers who received extra class credit for their participation.

Apparatus. A multidimensional compensatory tracking system, the Minnesota Tracking System (MTS), was used. For both experiments, the MTS was wired as a 2-dimensional system. Error information was displayed to *S*s in the form of a dot of light moving in 2 orthogonal dimensions on an 8 × 10 cm. oscilloscope screen. The *S*s controlled the movement of the dot in the vertical dimension by pushing and pulling a control rod with the right hand; pushing the rod moved the dot upward on the display, pulling moved the dot downward. Control in the horizontal dimension was accomplished by twisting a pistol-grip handle held in the left hand; twisting the control clockwise moved the dot to the right on the display, twisting counterclockwise moved the dot to the left. The range of travel of the right-hand push–pull control was 40 cm., corresponding to 11 cm. of vertical displacement of the dot, ±5½ cm. from the center of the display face. Control–display gain in the vertical dimension was 3.64:1.0. The range of travel of the left-hand twist control was 160°, corresponding to 11 cm. of horizontal displacement of the dot, ±5½ cm. from the center of the display face. Control–display gain in the horizontal dimension was 14.55:1.0. As is obvious, both controls were capable of moving the dot completely off the 8 × 10 cm. display face. The center of the display was indicated by the intersection of 2 hairlines; the centers of the ranges of both controls corresponded to this cross-hair-defined center. The MTS is described in detail by Dick (1970).

System forcing functions. Since the composition of the system forcing functions is critical to the experiments, it is necessary to describe them in some detail. In Experiment I, 2 pairs of forcing functions were used. Each pair consisted of 2 complex signals, 1 for the vertical dimension (A axis) and 1 for the horizontal dimension (B axis). All of the complex signals in both pairs were formed by combining a 5-cpm function with a 12-cpm function in the following manner. In the A axis, the complex signal was the resultant of the combination of a 5-cpm sine function with a 12-cpm sine function. The peak-to-peak amplitudes of the 2 functions were equal. In the B axis, the complex signal was the resultant of the combination of a 5-cpm sine function with a 12-cpm negative cosine function, again with the peak-to-peak amplitudes of the component functions equal. The 2 D conditions used in Experiment I were obtained by varying the interaxis phase relationship between the 2 complex signals to obtain the desired value of DC. A value of DC = 0.0 was the I stimulus, and DC = −.70 was the D stimulus. The result of these manipulations was one set of 5- and 12-cpm stimuli for the D experimental condition, and another set of 5- and 12-cpm stimuli for the I experimental condition. Since the period of the complex functions used in each axis is 60 sec., *S*s in both conditions saw 1 complete cycle of each stimulus signal within each trial. The only difference between the I and D conditions was the point in the cycle where the B axis began. In the D condition, a clockwise twist of the left-hand control was associated with pulling the right-hand control, and a counterclockwise twist was associated with

pu[illegible] the right-hand control. As the value of [illegible] for DC in the D condition indicates, the association was not perfect, but was a partial dependence. In the I condition there was no systematic relationship between the control movements required in the 2 dimensions.

The stimuli for Experiment II were generated in the same manner as in Experiment I, except that in addition to 2 sets of 5- and 12-cpm combinations, 2 sets of complex functions composed of 6- and 10-cpm sine functions were generated. The latter had the same dependence relationships as the 5- and 12-cpm signals, and the same peak-to-peak amplitudes. In Experiment II, the 2 sets (I and D) of 6- and 10-cpm signals were used as training stimuli, and the 2 sets of 5- and 12-cpm signals were used as test stimuli. In Experiment I, the same 2 sets of 5- and 12-cpm signals were used in both the training and test phases. All trials within each set of stimulus conditions were identical. All the forcing functions had peak-to-peak amplitudes sufficient to drive the dot ±4 cm. from the center of the display in both dimensions.

Procedure. The basic design of the experiments is shown in Table 1. Half of the *Ss* in each experiment were randomly assigned to the D training condition and half to the I training condition. At the end of the training phase, the training groups in Experiments I and II were randomly reassigned, half of each being tested on a task with the same (I or D) DC as in training and the other half on a dependence different from the training DC. This procedure yielded 4 groups in the test phases of Experiments I and II: I–I, D–D, I–D, and D–I.

The *Ss* were trained for 24 1-min. trials on the appropriate task, and then were shifted to their respective test tasks for 16 1-min. trials. Trials were presented in blocks of 8, the intertrial interval being 30 sec., and the interblock interval, 5 min. The *Ss* left the experimental room during the interblock interval while *E* prepared the next block of trials. The shift at transfer was made during the normal interblock interval, between the third and fourth blocks, and *Ss* were not informed of the change.

Instructions to *Ss* told them only that the experiment dealt with human motor skills, and that their task was to keep the dot centered on the screen by appropriate manipulations of the controls. The *Ss* were given a short pretraining familiarization with the action of the controls in the absence of a forcing function and then received their training and test trials in one experimental session. The *Ss* were aware that data were continuously recorded, but were not informed of the nature of the analyses to be performed. No feedback other than that present on the display during the course of a trial was given; nevertheless, *Ss* appeared to be well motivated and, without exception, were cooperative.

Measurement. The 2 response signals generated by *Ss*' manipulations of the controls and the forcing functions driving the dot in the 2 dimensions were simultaneously tape recorded for all trials. In the computer analyses, the 4 signals were simultaneously

TABLE 1

Design of Experiments I and II

Training condition	Experiment I	Experiment II
Dependence test condition *Ss*		
Dependence	3	6
Independence	3	6
Independence test condition *Ss*		
Dependence	3	6
Independence	3	6

sampled and converted to digital values at a 10-Hz. rate, and all analyses were performed on the digitized data. The sampling rate of 10 Hz. is sufficiently higher than the rates present in the sampled signals to assure good digital approximations to the original continuous signals. Using the digitized data, the following 4 measures of various aspects of *Ss*' responding and of system performance were calculated: rms error, DC, simultaneity index (SI), and an index of the goodness of fit of the multiplicative rule's prediction to the observed data. A brief description of the measures follows.

1. The rms error is the square root of the mean square tracking error over a trial. For each trial, each *S* has an rms error score for each dimension of the task. To derive error values from the recorded data required only that the computer calculate the difference between the amplitude of the forcing function in a given dimension and the amplitude of the response signal in that dimension at each sample point. These error values formed the basis for the rms error scores. The rms error is a sensitive, reliable measure of system performance (Bahrick, Fitts, & Briggs, 1957; Kelley, 1969). Decreasing rms error indicates increasingly good system performance. In addition, in a zero-order control system like the MTS, rms error is an index of the overall approximation of *Ss*' control position to the criterion control position.

2. The DC is the product-moment correlation of the rate-of-change in one axis with the rate-of-change in the other axis. The DC of the forcing functions is the criterion dependence required for perfect performance; the DC of *S*'s responding in the 2 axes is the dependent variable.

3. The SI is the ratio of simultaneous movements and nonmovements to the total number of sample periods, and is an index of output (response) time-sharing. Given a 4-fold table of movement (M) and nonmovement (NM) in Dimension A vs. M and NM in Dimension B, the M/A, M/B intersection is Cell A, the M/A, NM/B intersection is Cell B, the NM/A, NM/B intersection is Cell C, and the NM/A, M/B intersection is Cell D. For each 100-msec. sample interval during a trial, the computer decided whether *S* moved both controls, moved one control but not the other, or did not move either control. A movement was defined as a change in

the amplitude of the response signal of greater than 10 mv., which corresponds to a movement rate of .51 cm/sec on the vertical control and 3.16 degrees/sec on the horizontal control. For each sample interval the computer makes an entry in the appropriate cell of the table. At the end of a trial, SI is calculated as follows:

$$\mathrm{SI} = \frac{A + C}{A + B + C + D} \qquad [2]$$

The SI of the forcing functions used in Experiments I and II was approximately .80, which is the criterion value defined by the task. The SI of *S*s' responses is the dependent variable. It should be noted that because of the definition of very low rates as NMs, Cells B and D will be slightly inflated and as a consequence SI artifactually underestimates the actual simultaneity of the task and of *S*s. However, since both the task criterion SI and the SI of *S*s are underestimated to the same degree, the relation of the two is meaningful. (Cell C is also slightly inflated for the same reason, but since it appears in both the numerator and denominator of Equation 2 this does not affect the ratio.)

4. The method of determining the $P(\mathrm{TOT})$ scores from which to evaluate the goodness of fit of the multiplicative rule is somewhat out of the ordinary, and must be described in some detail. Bahrick et al. (1957) noted that the target size most sensitive to variations in system performance is one which encompasses approximately 68% of the error amplitude distribution. Since error in a tracking task varies as a function of practice, a given fixed target size will be more or less sensitive at various stages of practice. Therefore, for each *S*, the target size used was set equal to ±1 rms in each axis for each trial. This variable target was used to compute $P(\mathrm{TOT_a})$, $P(\mathrm{TOT_b})$, and $P(\mathrm{TOT_{ab}})$ for that trial in a post hoc analysis. This procedure results in fairly constant $P(\mathrm{TOT})$ scores for each individual axis and for the product of the single-axis scores, which is the prediction of the multiplicative rule. This technique of using post hoc target sizes based upon the dispersion of *S*'s own error amplitude distribution is due to Dick (1970). The goodness of fit index is then:

$$\text{Goodness of fit} = [P(\mathrm{TOT_a}) \times P(\mathrm{TOT_b}) - P(\mathrm{TOT_{ab}})] \times 100, \qquad [3]$$

where $P(\mathrm{TOT_a})$ and $P(\mathrm{TOT_b})$ are the proportions of TOT in each dimension, and $P(\mathrm{TOT_{ab}})$ is the proportion of TOT in both dimensions simultaneously. Assuming that mean tracking error is approximately zero and that the error amplitude distribution over a trial is approximately normal (Bahrick et al., 1957), then 1.0 rms ≐ 1.0 *SD* of the error amplitude distribution. From this it follows that the approximate range of the index is from −22% (maximum possible underestimation) to 10% (maximum possible overestimation).

In all the data analyses, only the middle 50 sec. of each trial were included. The first and last 5 sec. of each trial were dropped to avoid target acquisition artifacts and, at the end of a trial, the possibility of a synchronization error that would i[illegible] a few samples of posttrial noise from the tape containing the raw data. As a consequence, the values of DC defining the I and D conditions in the experiments were not precisely 0.00 and −.70 as originally planned, but varied slightly. In Experim[illegible]I, the D stimuli (both test and training) D[illegible].68, the I stimuli (test and training) DC = .0[illegible] In Experiment II, the D training stimuli D[illegible].66, the D test stimuli DC = −.71, the I train[illegible]timuli DC = .02, and the I test stimuli DC [illegible] The differences between the I and the D D[illegible]ppear to be negligible.

Results

This section focuses primarily on the results of Experiment II. With respect to the task dependence variable, the 2 experiments are replicates, so the results of Experiment I will be introduced only as necessary to support the Experiment II data. In general, the Experiment I results parallel those of Experiment II.

System performance. The measure of system performance, rms error, showed clear change as a function of practice and as a function of the change in dependence at the shift between Blocks 3 and 4. It decreased over trials of training, approaching asymptotic levels in the second and third blocks of training in both experiments. The I group tended to have slightly higher error through the training trials. In Experiment II, at the shift between Blocks 3 and 4, all 4 test groups showed an increase in system error. This in itself is not surprising, since the test-task forcing functions contained both higher and lower rates than the training forcing functions. However, the shift groups (I–D and D–I) showed greater increases from Block 3 to Block 4 than did the nonshift groups. Table 2 shows the ratio of rms error in Block 4 over rms error in Block 3 for the 4 test groups. Each entry in the table is the mean rms error in Block 4 of one of the test groups divided by the Block 3 mean rms error for the 6 *S*s of that test group. All of the ratios are greater than 1.0, indicating that system error increased as a function of the task change in all 4 groups. Also apparent in the table is a pattern reflecting

fects of the changed dependence for shift groups. They show greater increase in rms error than the nonshift groups. There was an increase in system error due to the change in task dependence over and above that due to the change in the intra-axis characteristics of the task. The Experiment I data showed the same general pattern. With the exception of 1 *S* in the I–I group, the nonshift groups improved (i.e., ratios < 1.0), which is reasonable since in Experiment I their tasks were identical in training and test. The shift groups, on the other hand, showed increases in system error from training to test. These increases can only be due to the change in dependence, since the intra-axis characteristics of the training and test tasks were identical in Experiment I. This result was consistent at the individual *S* level. In Experiment I at this level there are 24 ratios (12 *S*s $\times$ 2 variables), and of these 24 ratios only 3 depart from the pattern described. Of the shift *S*s' ratios, 11 of 12 were greater than 1.0; of the nonshift ratios, 10 of 12 were less than 1.0.

A transfer analysis comparing shift groups with nonshift controls, using Briggs' (1969) Equation 1, was performed on the Block 4 rms error data.[3] This transfer index is:

$$\frac{\text{Control group} - \text{Experimental group}}{\text{Control group}} \times 100. \quad [4]$$

For rms error, which decreases with improved system performance, the index varies from 100% to minus infinity. Table 3 presents the transfer indices calculated from the Block 4 mean rms error data of Experiment II. All of the indices have negative signs, indicating greater rms error in the shift groups. In addition, the I–D group showed greater negative transfer than the D–I group. Again, the Experiment I data replicate the pattern of Table 3, showing general negative transfer (with the exception of 1 aberrant *S*), and differentiating the D–I from the I–D group, in spite of the aberrant *S*. The transfer indices, together with the ratios of Table 2, show that there was a detrimental effect on system performance due to the change in task dependence at transfer.

Fit of the multiplicative rule. The multiplicative rule generally predicted observed *P*(TOT) very well in both experiments. The range of the block mean goodness of fit indices varied over less than 10% of the total possible range of the index. The rule generally underestimated observed *P*(TOT), and at the block mean level of analysis appeared to segregate groups by task dependence, the D groups being underestimated more than the I groups in both training and test blocks. Both experiments produced the same patterns of results in these respects.

However, the goodness of fit data were not as regular as the block means imply. Figure 1 shows the goodness of fit index over the last 8 training trials and the first

[3] The terms in the numerator of Equation 4 are reversed from Briggs (1969) so as to yield negative transfer indices in those cases where the experimental group error was greater than the control group error. Relative to a (hypothetical) control group receiving no training prior to the test trials, of course, all test groups showed positive transfer.

TABLE 2

RATIO OF ROOT MEAN SQUARE (RMS) ERROR IN BLOCK 4 TO RMS ERROR IN BLOCK 3 FOR 4 TRAINING-TEST CONDITIONS OF EXPERIMENT II

Condition	Dimension A	Dimension B
Independent–independent	1.02	1.02
Dependent–dependent	1.03	1.01
Independent–dependent	1.06	1.12
Dependent–independent	1.06	1.29

TABLE 3

TRANSFER INDICES COMPARING ROOT MEAN SQUARE (RMS) ERROR OF SHIFT GROUPS WITH RMS ERROR OF NONSHIFT CONTROL GROUPS: EXPERIMENT II, BLOCK 4 MEAN DATA

Condition	Dimension A	Dimension B
Independent–dependent	−18.7%	−16.4%
Dependent–independent	− 7.4%	− 6.3%

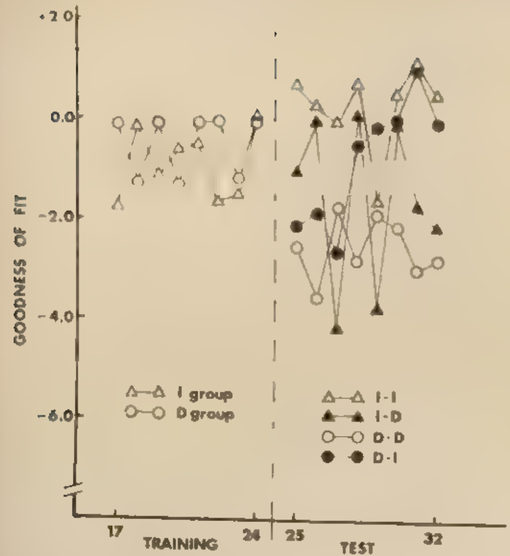


FIGURE 1. Goodness of fit of the multiplicative rule over the last 8 training trials (Block 3) and first 8 test trials (Block 4) of Experiment II. (Abbreviations: I = independent, D = dependent.)

8 test trials of Experiment II. As is obvious, the fit of the rule to the data was neither regular nor clearly related to task dependence. (Block 3 of Experiment II was the only place where the D and I *Ss* were not separated at the block mean level of analysis.) There is considerable overlap of the trial mean goodness of fit in the various test groups, and it is not the case that the fit of the multiplicative rule is unequivocally determined by the task dependence, where dependence is defined in terms of DC. Again, the Experiment I data agree.

Behavioral dependence. The interaxis response dependence displayed by *Ss*, indexed by DC, is shown over the 40 trials of Experiment II in Figure 2. (The DC data are in the form of logarithmic transformations of *r*. This transformation was performed for purposes of analyses not germane to the present discussion, and does not affect the results reported here.) As the figure shows, *Ss*' interaxis control movement dependence changed as a function of practice, and the change is in the direction of the task-required dependence. The training-trial means approach the criterion dependence in both conditions, the D *Ss* moving in a more dependent manner and the I *Ss* moving more independently, but in neither condition did they reach the criterion in training.

The change in DC at the shift appears to be quite regular. The D–D group increases DC slightly, corresponding to the slight increase in task dependence, and the I–I *Ss* change DC slightly in a positive direction, again corresponding to a small change in the task dependence. The shift groups (I–D and D–I) also change dependence rapidly. The I–D group, while quickly changing dependence at the shift, does not change to the value of the D–D control group, but falls short by a small margin. In Block 3, the D–D group block mean was at 79.7% of the criterion value, while in Block 4 it was at 79.5% of the criterion. That is, there was no substantial change in the D–D *Ss*' relationship to the task criterion DC as a result of the change from a dependent 6- and 10-cpm task to an equally dependent 5- and 12-cpm task. The I–D group, however, reached only 70.5% of the criterion dependence in the first test block. This pattern of results was replicated in Experiment I, where the D–D *Ss* were at 94.0% of criterion in Block 3 and 95.2% in Block 4, while the I–D group reached only 73.5% of criterion in Block 4. The I–D trial means are consistently below those of the D–D *Ss* through Block 4 in both experiments. A 2-way analysis of variance with repeated measures on one

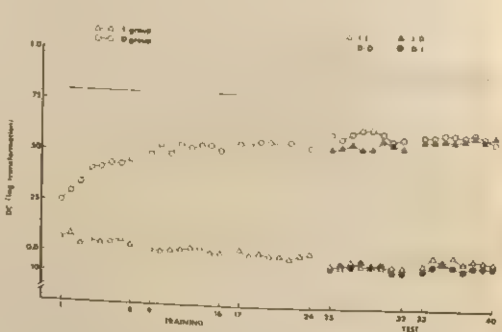


FIGURE 2. Mean dependence coefficient (DC) over the 40 training trials of Experiment II. (The horizontal bars over the data points indicate the task criteria DC for the dependent [D] group; the independent [I] group criteria were .02 in training and .04 in test.)

ctor (trials) was used to assess the tatistical significance of the effect of training dependence (I or D) on the DC of *S*s' responding on a D task in Block 4 of Experiment II. Neither the training dependence main effect, F (1, 10) = 2.18, the trials main effect, F (7, 70) = .60, nor the Training Dependence × Trials interaction, F (7, 70) = .67, was statistically significant. However, the replicability of the phenomenon renders the nonsignificant outcome of a statistical inference based on one set of data less than compelling.[4]

The shift in dependence from D to I showed no effects like those just described. The D–I *S*s shifted directly to the new dependence at transfer, and were indistinguishable from the I–I controls at both the block and trial mean levels of analysis. This result also replicated in Experiment I and in the partial replication mentioned in Footnote 4. In general, the results of the DC analyses show that while training on a D task has no discernible effects on the dependence displayed on a subsequent I task, the reverse is not true. Training on an I task is not equivalent to training on a D task with respect to subsequent responding on a D test task. This asymmetry is reflected in the rms error transfer indices in which the I–D groups show greater negative transfer than the D–I groups in both experiments. The greater negative transfer in the I–D groups' system error may be due to the failure to move in as dependent a manner as the D–D control groups in Block 4. This is not to say that there is *no* commonality of what is learned in the 2 tasks; it is obvious that *S*s in both training conditions are learning very similar skills, but they are also learning something different: a relationship in the D condition between particular movements of one control

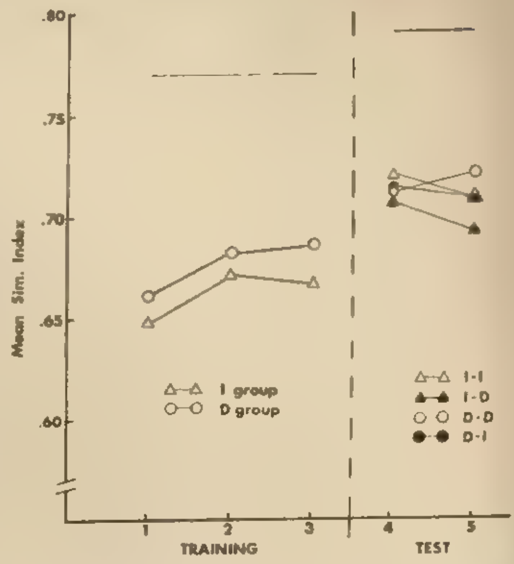


FIGURE 3. Simultaneity index (SI) over blocks of 8 training and test trials of Experiment II. (Horizontal bars over data points indicate training and test SI criteria. Abbreviations: I = independent, D = dependent.)

and the other control which did not characterize the I condition.

The SI is shown over blocks of trials of Experiment II in Figure 3. Note the expanded scale on the vertical axis of the figure. The total possible range of the index is 0.0–1.0; the vertical axis includes only 20% of this range. In terms of the total range of the index, SI changed very little as a function of practice. Although the task SI was not a manipulated variable in the experiments, it did change slightly in Experiment II, increasing from .77 to .79 with the shift in tasks at transfer. The *S*s' SI changed abruptly at the shift, paralleling the change in criterion SI. In training there appeared to be a segregation of the groups by task dependence, but it did not remain consistent in the test blocks, even for the nonshift groups. The Experiment I data replicated the Experiment II results in that in training, D *S*s exhibited a slightly higher SI than I *S*s, while in test the groups were not differentiated according to task dependence. Both training conditions within each experiment had the same criterion SI, and the author has no

---

[4] It is worth noting that a third partial replication, in which 2 groups of 3 *S*s each were trained on 6- and 10-cpm I and D tasks and then shifted to the other (D or I) dependence, yielded results (compared to extrapolations of training data) showing the same patterns of transfer asymmetry as the 2 full experiments reported here. The fact that the patterns replicated 3 times is sufficiently strong evidence to show that they did not occur by chance.

satisfactory explanation for the separation in training.

The SI does not appear to be clearly related to the fit of the multiplicative rule to Ss' data. While the SI data indicated a consistent differentiation between dependence groups in training, the multiplicative rule's fit to their $P(TOT)$ data varied considerably. Changes in the goodness of fit of the rule did not reflect changes in response time-sharing patterns.

Discussion

It was noted in the Results section that the fit of the multiplicative rule was not determined by task dependence, but there was no particular reason to suppose that it would be. The rule has been taken to be an indicant of *response* independence, an index of the degree to which Ss' responses in one component of the task are independent of those in the other component. It is clear by now that the rule is not directly related to response dependence whether dependence is defined as movement simultaneity (SI) or as the correlation between the rates of movement of the 2 controls (DC). The training data of Experiment II provide sufficient reason to reject both interpretations of the rule: While the rule's predictions were becoming more accurate over training blocks, the D and I group means becoming almost identical in this respect, both DC and SI clearly differentiated Ss in the 2 conditions. As a consequence, it must be concluded that the multiplicative rule for TOT scores, while a powerful empirical relationship between system performance component measures, does not reflect an underlying behavioral independence. The SI training data suggest that a D track, in which there is statistical predictability from one control to the other, tends to encourage response time-sharing, but the test data for the I I and D D groups do not bear this out.

The data indicate that Ss can and do impose a dependent structure on their responding when the task situation requires it. On the other hand, if the task requires independent movements, Ss can supply them. Given trials of practice, Ss approach the dependence defined by the task as criterion. Over and above this, it appears that task dependence may be part of what Ss learn in the practice trials. The relationship of the rms error transfer indices to the asymmetry of the DC shifts suggests that the I–D group had not learned everyt ng there was to know about responding on e D test task despite the intraaxis similarit f the 2 conditions in Experiment I. The rep tion of this pattern in Experiment II in face of a concomitant change in the intra characteristics of the task implies that I–D Ss must learn the dependent res patterning required by the new task. D–I Ss changed dependence immediately transfer, implying that they had acquired skill necessary to pattern their movem independently prior to the test trials, tho one should note that their rms error scores creased from training to test. It is not cert that this skill was acquired in the context the experimental training; it is possible that was acquired preexperimentally. There ma be population stereotypes regarding the inte limb patterning of movements in multidimen sional tasks. Research to explore this question is currently in progress.

The present research is the first in which question of human tracking behavior has bee explicitly studied in the light of the behavior requirements of the task. The view taken ha been that tracking is a behavior reproductio task. The S's job, from E's point of view, to reproduce the RRP defined by the syster forcing functions and the electromechanic characteristics of the control system. In ord to study behavioral dependence, it is necessary to specify the dependence required by the task and to evaluate Ss' responding with reference to this criterion. This approach would have led Dick (1970) to specify the phi correlation required by his task, as was done for SI and DC in the present research, so that the shifts in time-sharing he observed would have been interpretable. Similarly, Adams and Webber (1961), in attempting to ascertain the relationship between behavioral independence and the multiplicative rule, cross correlated error records. They found time lagged correlations that were reliably nonzero, and concluded that there was underlying behavioral dependence in spite of the fit of the multiplicative rule's predictions. However, Adams and Webber give no information regarding the task requirements in this respect beyond saying that they used "input signals that were . . . unsystematic between component tasks [p. 203]." Whether "unsystematic" means that the cross-correlation function relating the forcing functions in the 2 axes of their task did not depart from zero is unknown. The information necessary to interpret the data

reported is unavailable, and one cannot know just what the *Ss* were doing relative to the task requirements.

The research reported in this article supports the utility of the view of tracking, for measurement purposes, as a behavior-reproduction task. This view assumes that one's task as an experimental psychologist is to produce research answers to theoretical questions. Given that theories in psychology make predictions about behavior, it is logically necessary to have a research and measurement methodology capable of generating answers phrased in terms of behavior. It follows that the experimental task used must be described in terms of behavior, and the measures used must refer unambiguously to behavior. Thus the view of tracking as a behavior-reproduction task is not a theory of motor behavior, but is the basis for a research methodology appropriate to the study of that behavior.

REFERENCES

Adams, J. A., & Webber, C. E. The organization of component response error events in two-dimensional visual tracking. *Journal of Experimental Psychology*, 1961, **61**, 200–212.

Bahrick, H. P., Fitts, P. M., & Briggs, G. E. Learning curves—Facts or artifacts? *Psychological Bulletin*, 1957, **54**, 256–268.

Bilodeau, E. A. Variations in knowledge of component performance and its effect upon part-part and part-whole relations. *Journal of Experimental Psychology*, 1955, **50**, 215–224.

Bilodeau, E. A. Patterns of internal consistency in multipart skilled performances. *American Journal of Psychology*, 1957, **70**, 550–559.

Bilodeau, E. A., & Bilodeau, I. McD. The contribution of component activities to the total psychomotor task. *Journal of Experimental Psychology*, 1954, **47**, 37–46.

Briggs, G. E. Transfer of training. In E. A. Bilodeau (Ed.), *Principles of skill acquisition.* New York: Academic Press, 1969.

Dick, R. A. Independence of component tasks in a multidimensional tracking task. Unpublished doctoral dissertation, University of Minnesota, 1970.

Ellson, D. G. *The independence of tracking in two and three dimensions with the G. E. Pedestal Sight.* (Rep. TSEAA-694-2G), USAAF Air Materiel Command, 1947.

Fitts, P. M., & Posner, M. I. *Human performance.* Belmont, Calif.: Brooks/Cole, 1967.

Hoppe, R. B. Inter-control movement dependence in a two-dimensional compensatory tracking task. Unpublished doctoral dissertation, University of Minnesota, 1972.

Kay, H. The development of motor skills from birth to adolescence. In E. A. Bilodeau (Ed.), *Principles of skill acquisition.* New York: Academic Press, 1969.

Kelley, C. R. The measurement of tracking proficiency. *Human Factors*, 1969, **11**, 43–64.

O'Donnell, P. A. *Motor and haptic learning.* San Rafael, Calif.: Dimensions, 1969.

(Received February 27, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 226-233

CROSS-TASK VALIDATION OF FUNCTIONAL MEASUREMENT USING JUDGMENTS OF TOTAL MAGNITUDE[1]

NORMAN H. ANDERSON[2]

University of California, San Diego

The *Ss* rated total magnitude of 5 kinds of stimulus pairs: length length, area area, length area, length weight, and area weight. An adding model fit the data in each case, providing a within-task test of consistency. Functional measurement scale values for each dimension were linearly related across tasks, providing a between-task test of consistency. These 2 kinds of consistency tests validate the rating response as an interval scale of subjective magnitude, lending further support to previous conclusions that magnitude estimation is biased and invalid.

A perennial idea over the last 4 decades of work on psychophysical scaling has been the use of one or another algebraic model as a basis for measurement (see Anderson, in press-a, in press-b). Ratio models for fractionation and adding models for stimulus summation are examples. Such models impose a criterion of internal consistency on the data thereby constituting a validational base for the scales. Psychophysical data, however, have not generally fulfilled such criteria.

Magnitude estimation deserves particular note because of its widespread use. It is well known, however, that magnitude estimation data have not generally satisfied criteria of internal consistency (see, e.g., Anderson, 1970b, in press-a; Garner & Creelman, 1967; Poulton, 1968). Without such a validational base, magnitude estimation can be considered no more than an ordinal scale.

Recent developments show more promise. Functional measurement methodology (Anderson, 1970b), and related techniques developed by other investigators, begun to yield results that do show ternal consistency. This approach r heavily on the use of algebraic models psychophysical integration. The inte tion model serves as a scaling frame which to construct interval scales for stin lus and for response. Scaling then beco an organic part of substantive inquiry, a separate methodological preliminary.

An important outcome of this work is the support for ordinary rating methods. With modest experimental care, ratings can yield responses that are on an interval scale. Magnitude estimation, in contrast, seems to be biased and nonlinear (e.g., Anderson, 1972a; Weiss, 1972), a conclusion that is also suggested by the work of the Oregonians (e.g., Beck & Shaw, 1968; Curtis, Attneave, & Harrington, 1968; Curtis & Fox, 1969; Fagot & Stewart, 1969).

This report takes up an interesting task introduced by Feldman and Baird (1971) who asked *Ss* to judge "total intensity" of a sound plus a light. This kind of stimulus integration seems to call for an adding model. The adding model constitutes a criterion of internal consistency, and its success would allow construction of interval scales of the stimuli using functional measurement.

The total magnitude task also provides a criterion of cross-task consistency. Since *Ss* seem able to add stimuli from quite different modalities, diverse physical dimensions can be scaled under directly comparable conditions. Cross-task consistency

[1] This work was supported by National Science Foundation Grant GB-21028, and facilitated by grants from the National Institute of Mental Health to the Center for Human Information Processing, University of California, San Diego. I wish to thank Holly Halvorson for research assistance, Michael Birnbaum for helpful comments on the manuscript, and Dwight Curtis and Stanley Rule for making available data and information cited in the text, and for comments on a draft of this paper.

[2] Requests for reprints should be sent to Norman H. Anderson, Department of Psychology, University of California, La Jolla, California 92037.

(Weiss & Anderson, 1969) requires that the scales for any one dimension should be the same across the different pairs in which it occurs. The present article gives an initial application of functional measurement methodology to this total magnitude task.

Method

Stimuli

The 3 sets of stimuli were length, area, and weight, with 5 stimulus levels in each set. The *lengths* ranged 4–20 cm. in 4-cm. steps. They were constructed of black drafting tape, 2-mm. wide, stuck to a white tagboard approximately 8 × 28 cm. The *areas* were circles with diameters ranging 2–6 cm. in 1-cm. steps. They were constructed of black paper, pasted on white tagboard approximately 10 × 14 cm. The *weights* ranged 200–600 gm. in 100-gm. steps. They were not visible, and were lifted by pulling down on a handle-pulley arrangement used in previous work (Anderson, 1972a).

The *S*s judged stimulus pairs, of which there were 5 types. Three were heteromodal (length–area, length–weight, and area–weight) and 2 were homomodal (length–length and area–area). The weight–weight condition was omitted as it would have involved serial rather than simultaneous presentation. The 3 heteromodal pair types were run in complete 5 × 5 factorial designs. Owing to an error, the design for the homomodal pairs was triangular, not a complete factorial, but that did not affect the results.

A small and a large stimulus end anchor, constructed from corresponding single stimuli, were used for each type of pair. The small and large lengths were 2 and 25 cm., the small and large circles were 1 and 8 cm., and the small and large weights were 100 and 800 gm. These end anchors had their usual purposes, including that of reducing end effects in the rating scale (e.g., Anderson, in press-a, in press-b; Eriksen & Hake, 1955, 1957).

Procedure

The *S* was instructed to judge the "total intensity" of each pair of stimuli. Ten practice trials covering the stimulus range were given for each type of pair. The first 2 were the small and large anchors, presented with instructions that they should be called 1 and 20, respectively, with the other pairs to be given intermediate numbers according to their total magnitude.

Pairs involving length and area were presented by placing the 2 stimuli in front of *S*. Pairs involving weight were presented by having *S* lift the weight while looking at the other stimulus.

One replication of the 5 stimulus designs was run in each of 3 sessions. Practice was given only in the initial session, but the 2 anchors were presented anew before each later replication. In each replication, the length–area pairs were presented first, followed in order by length–weight, area–weight, length–length, and area–area. Responses for each stimulus pair were recorded on an index card. The deck of cards for each stimulus type was shuffled for each replication for each *S* to balance order effects. The *S*s were 20 students or similar members of the university community who were paid $5.64 for the 3 sessions.

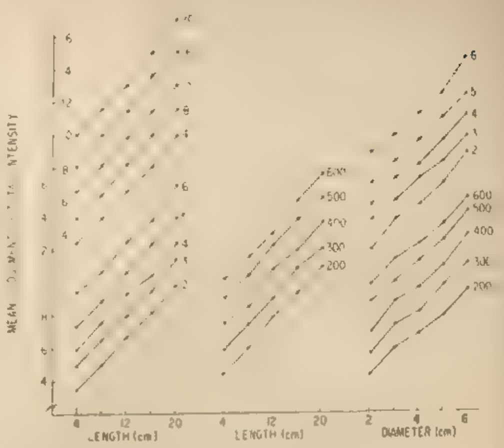


FIGURE 1. Parallelism test of integration model. (Vertical axis represents mean judgment of total magnitude. Lower panels for heteromodal stimulus pairs; upper panels for homomodal stimulus pairs.)

Results

Integration Model

The mean judgments of the stimulus pairs are presented in Figure 1. Each lower panel shows one heteromodal type, plotted in the format of the 5 × 5 design. The values of one design factor are given on the horizontal, and the values of the other design factor are listed by each curve. The 2 upper panels show the homomodal data plotted analogously. For comparison, the upper panels are plotted as though a complete factorial rather than a triangular design had been used; each point connected by a dotted line is identical with a corresponding lateral point connected by a solid line.

The key property of these data is parallelism. The simple adding model predicts that each set of curves should be parallel, and that is essentially true. A rigorous test of parallelism is available

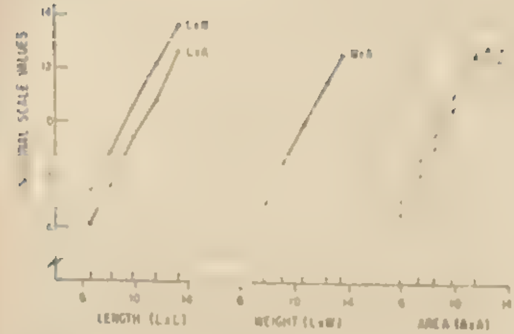


FIGURE 2. Linearity test of cross-task consistency. (Each curve represents the functional scale from one type of stimulus pair plotted against the functional scale from another type of stimulus pair.)

from analysis of variance; the ordinary Row × Column interaction tests for significant discrepancies from parallelism. This interaction did not approach significance in any test, F (16, 304) = 1.49, .79, and .78, in the 3 respective panels. The 3-way interactions with replications were also nonsignificant. Similar tests are possible for the 2 triangular designs (Anderson, 1972b), but were not made in view of the evident parallelism in the graphs.

The simplicity of this analysis is notable. Simply plotting the raw ratings gives an immediate graphical test of the integration model. The parallelism property jointly validates the integration model and the response scale since in general both must hold for parallelism to obtain. Functional scales of the subjective stimulus values are also available. These are considered next.

Functional Scales

Since the model has passed the parallelism test, subjective scales of the stimuli can readily be obtained. According to functional measurement theory, these are just the marginal means of the factorial design.

Scaling can be illustrated for the Length × Area data in the lower left panel of Figure 1. The vertical spacing of the curves constitutes an interval scale of the subjective values of the areas. The curves for circles of 2- and 3-cm. diameter appear to be closer together than the curves for circles of 5- and 6-cm. diamete[illegible] Thus, subjective circle value appears t[illegible] ow a small positive acceleration as a [illegible]ction of physical diameter. In this sa[illegible] plot, the mean curve yields a scale of f[illegible]ional length. It is almost linear in [illegible]s of physical length.

Functional scales from the oth[illegible] fac-torial designs are obtained in [illegible] same way. Scale values from the t[illegible]ular designs can be obtained analogou[illegible]

Cross-Task Validation

Three functional scales of sub[illegible]tive length are available since the length [illegible]timuli were used in 3 types of pairs. [illegible] the same subjective values function in each of the 3 integration tasks, each l[illegible]ngth scale should be a linear function of each other length scale.

This test of cross-task consisten[illegible] is shown in Figure 2. In the left pan[illegible], the functional scale of length from the L[illegible]ngth × Length design is plotted on the hori-zontal axis. The other 2 functional scales of length, from the Length × Area and Length × Weight designs, are plotted on the vertical. Both curves are essentially linear. Thus, the subjective values of length are indeed the same, up to a linear transformation, across all 3 integration tasks.

The center panel plots subjective weight from the Area × Weight design against subjective weight from the Length × Weight design. The right panel shows subjective area from the Length × Area and Area × Weight designs as a function of subjective area from the Area × Area design. All these curves appear linear. In general, therefore, the same subjective values appear to function in each task.

Psychophysical Law

Once the subjective stimulus values have been obtained, the psychophysical law can be obtained very simply. It corresponds to the graph of these subjective values as a function of the objective physical values. These graphs are shown in Figure 3.

The 3 psychophysical functions for length (left panel of Figure 3) are all

essentially linear. This result agrees with most previous work on length.

The center panel plots the 2 psychophysical functions for heaviness. Both show a slight downward convexity. This shape, which would correspond to an exponent slightly less than 1 in a power function, agrees with previous work in this program (Anderson, 1970a, 1972a). In contrast, magnitude estimation yields an exponent around 1.45 (Stevens & Galanter, 1957). This disagreement is considered further in the Discussion section.

The 3 curves in the right panel of Figure 3 show the psychophysical function for area. These appear to have some upward convexity, as might be expected since physical area would itself be convex upwards as a function of circle diameter.

It should be noted, however, that a linear function would give a reasonable fit to these curves for area, suggesting a tendency to rely on the diameter, perimeter, or other linear aspect as a measure of the magnitude of the circle. Since such a tendency might be potentiated by the present array of tasks, these scales for area may be task specific. More appropriate scales of area might be obtained by restricting to designs with areas only and by emphasizing area judgments in the instructions. Thus, circles could be paired with triangles, crosses, or disks. This procedure could also provide subjective scales for irregular figures by applying a 2-stage model (Anderson, in press-a; Anderson & Weiss, 1971).

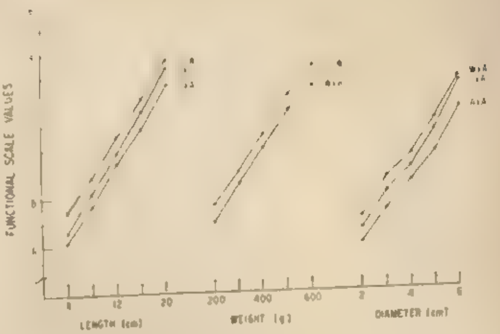


FIGURE 3. Psychophysical functions: Functional stimulus scales are plotted against the physical values.

Related Work on Weight Adding

Curtis and Fox (1969) applied an adding model to judgments of sums of 2 lifted weights. They reported a good fit for their rating data, in agreement with tests of an averaging model for lifted weights (Anderson, 1967; Anderson & Jacobson, 1968).

However, these 2 integration tasks, summing and averaging, have produced a seeming disagreement about the psychophysical function for lifted weight. The relevant reports all agree that the psychophysical function is convex downwards, which would correspond to an exponent less than 1 in a power function. However, Curtis and Fox (1969) reported an exponent around .65, whereas Anderson (1972a) claimed a near linear function by visual inspection. This seeming disagreement prompted further analysis of the previous data, and 3 main points need mention.

Data analysis. First, S. J. Rule (personal communication, March 16, 1973) fit a power function to Anderson's (1972a, Figure 3) data, obtaining an exponent of .69. This value is close to that of Curtis and Fox (1969). With a fixed response range, of course, the difference in stimulus ranges (10–218 gm. in Curtis and Fox, 200–600 gm. in Anderson) would itself produce some difference in the exponent of a power function fit (Poulton, 1968). It may be worth adding that a logarithmic function remains invariant, changing only in its base to adapt different stimulus ranges to a fixed response range.

Second, the present author reanalyzed the category rating data of Curtis and Fox (1969) using functional measurement to test the adding model. Analysis of variance for triangular designs (Anderson, 1972b) was applied separately to each single subject and to the group as a whole. These analyses supported the adding model fairly well. For the group data, deviations from additivity were not significant tested against the *S*s × Pairs interaction, F (20, 243) = .97. Of the 10 single *S*s, 2 or 3 did show significant discrepancies, tested against individual within-cell variability, but these did not seem to be overly serious.

For category ratings, therefore, the data of Curtis and Fox (1969) agree well with results on stimulus integration obtained by the author in the present and previous reports.

Third, the present author also reanalyzed the magnitude estimation data obtained by Curtis and Fox (1969). In the individual analyses, 7 of the 11 *Ss* showed marked discrepancies from the adding model, with 6 of the 7 *F* ratios significant beyond the .001 level. If the adding model used by Curtis and Fox is correct, magnitude estimation cannot be valid.

In the group magnitude estimation data, these individual discrepancies from the adding model averaged out. Indeed, the adding model fit the group means quite well, F (19, 260) = .34, for deviations from additivity. The group geometric means were not much different from the arithmetic means, in part perhaps because of the choice of standard and modulus in this particular experiment. This example illustrates the risk in the standard procedure of averaging magnitude estimation data across *Ss*.

Statistical analysis. The statistical analysis used by Curtis and Fox (1969) has certain limitations that need to be pointed out. The basic limitation is that it does not provide an adequate test of the adding model. In slightly changed notation, their 2-stage model would read

$$R = a(S_1^k + S_2^k)^m + b,$$

where S_1 and S_2 are the physical values of the 2 stimuli, R is the observed response, k and m are power function exponents, and a and b are constants.

They fitted this equation using the standard, nonlinear least squares analysis to estimate the exponent parameters in the above equation. Such analyses also purport to give certain statistical reliability information, including standard errors of the parameters, together with an overall F test of deviations from the model. This reliability information, unfortunately, has little value because it rests on the assumption that the fitted model is correct. In this analysis, deviations from additivity, from the power functions, or both, go into the error term. Incorrectness in the model thus tends to obscure itself because it inflates the error term. To the best of the present author's understanding, this kind of analysis needs to be restricted to the single *S*, and requires an error term based on within-cell variability for each *S*.

The functional measurement procedure may be useful in further study of the 2-stage model. It is more general than the procedure used by Curtis and Fox (1969) because it requires no assumption about the psychophysical stimulus function. Indeed, no physical metric is needed for the stimuli which may have merely nominal values. General response transformation is also allowable (Anderson, in press-a, Section II, A, 6; Bogartz & Wackwitz, 1971; Weiss, 1973). Thus, a rigorous test of the basic model is feasible while allowing for general input and output functions.

It should be emphasized that the present analyses have reached the same conclusions as Curtis and Fox (1969), excepting only the present finding of substantial discrepancies in the individual data on magnitude estimation. The limitations in their statistical technique are not of present importance, therefore, but they need mention because similar techniques have been employed in related reports. These statistical tests of goodness of fit are important because the theoretical interpretation depends on whether or not the basic algebraic model is correct.

Other Related Work

The heteromodal integration task was introduced by Feldman and Baird (1971) who tested an adding model (arithmetic mean) and a multiplying model (geometric mean) for judgments of total magnitude of stimulus pairs that consisted of a sound and a light. Their graphs suggest that the adding model did better, but they do not provide a test of goodness of fit. The functional measurement procedure would give a much simpler and more powerful analysis. The adding model predicts parallelism, while the multiplying model pre-

dicts a bilinear form (Anderson, in press-a, Section II, B, 6; in press-b, Section 7.12).

Dawson (1971) studied sum and difference judgments of circular area and of loudness, and Marks and Cain (1972) studied difference judgments of circular area, heaviness, and roughness. Both reports concluded that there were systematic discrepancies from the sum and difference models. However, these discrepancies may merely result from bias in the magnitude estimation response.

One other report on sum judgments should also be noted. Five experiments on length adding in a thoughtful article by Kreuger (1970) all found systematic overestimation of the combined length of 2 or 4 lines, relative to judgments of single line controls. Although Kreuger concluded that the integration procedure itself introduced a bias into the length production response, a simple adding model may be able to account for his data. To a large extent, though perhaps not completely, the overestimation could reflect a constant error introduced by the combining process. That would be consistent with the rough tendency toward constant overestimation across different lengths, and with the greater overestimation for 4 lines than for 2 lines.

Discussion

Integration Theory

The present data support the adding model for psychophysical integration in all 5 cases. The simple parallelism test of Figure 1 accomplishes 3 goals simultaneously (Anderson, 1970a, Figure 1): (*a*) It validates the integration model (psychological law), (*b*) it validates the response scale, and (*c*) it yields interval scales of subjective stimulus value.

Numerical response measures, such as ratings, may well fail to be interval scales. But if the ratings were not an interval scale, a correct model would fail the parallelism test. That test, therefore, provides a joint validation of the model and of the response scale. The stimulus scales derive from the model as already shown.

The basic element in this functional logic is the integration model. It is the structural frame on which response and stimuli are scaled. The simple adding task may seem trivial in itself, but it has an important theoretical role to play (Weiss & Anderson, 1969, 1972).

Magnitude Estimation Seems to be Invalid

The present results imply that magnitude estimation is biased and invalid. It has been repeatedly emphasized (e.g., Stevens, 1957, 1966, 1971) that ordinary ratings and magnitude estimations are nonlinearly related for prothetic dimensions such as were used here. This point is well illustrated by the present psychophysical function for lifted weight which would have an exponent less than 1 in a power function fit. In contrast, magnitude estimation yields exponents around 1.45 (Stevens & Galanter, 1957; see also Stevens & Rubin, 1970). Being nonlinearly related, ratings and magnitude estimates cannot both be interval scales of the same thing. At least one is invalid.

The present rating data have satisfied 2 validity criteria. The parallelism test of Figure 1 demonstrates within-task consistency, providing one basis for concluding that the rating response is an interval scale. The linearity test of Figure 2 demonstrates cross-task consistency, thus giving generality to the concept of stimulus scale value. Since these results support the validity of the rating response, they imply that magnitude estimation is invalid.

This experiment does not stand alone. Three other lines of evidence support the same conclusion. The first is a program of work in the author's laboratory that has provided substantial support for simple algebraic models of psychophysical integration. In particular, Anderson (1972a) showed within-task and cross-task consistency of heaviness scales obtained from weight averaging and from the size-weight illusion. Also, Weiss (1972) made a direct comparison in grayness averaging and found that ratings satisfied the parallelism prediction and magnitude estimation did not.

The second line of evidence comes from other work on psychophysical integration. Studies by Curtis and his colleagues have generally found that magnitude estimations do not fit sum and difference models, unless perhaps they are nonlinearly transformed. These same studies have also generally appeared favorable to ratings though correct tests of fit have been lacking as noted above. Fundamental work by Birnbaum and Veit

(in press) can also be most simply interpreted to support ratings as opposed to magnitude estimations.

The third and final line of evidence comes, curiously enough, from social psychology. Using social stimuli, Stevens (1966) has found the same nonlinear relation between ratings and magnitude estimations that is obtained with prothetic psychophysical dimensions. He concluded that ratings show the same general bias in both cases. However, a long-extended program of work by Anderson (1962, in press-b) has found that ratings also satisfy a class of algebraic integration models in the social field. Basic work in decision making by Shanteau (e.g., 1970, 1972) has also supported the functional measurement approach.

The social data have particular interest. Unlike most psychophysical dimensions, many social dimensions have both positive and negative values. Such data make it plausible to rule out geometric averaging models and analogous models that might seem compatible with magnitude estimation data (Anderson, 1972a; Birnbaum & Veit, in press; Weiss, 1972).

Integration and Valuation

Heteromodal integration raises an interesting question about process. Since length and weight are qualitatively different sensations, how can they be added together? To handle this conceptually, it would seem reasonable, if indeed not necessary, to suppose that the different stimuli are changed into some common currency for the demands of the immediate task. Sensations are not themselves added, but rather their processed values.

Integration theory handles this matter in terms of a valuation operation. Any task defines a judgment dimension, implicitly or explicitly. The valuation operation produces the scale value of the stimulus, and its weight, in relation to this judgment dimension.

But this view requires that the scale value depend on the integration task because that controls the valuation operation. It would be premature, therefore, to claim that scale values from any one task will measure the magnitude of a basic sensation. Each task imposes its own valuation operation, and these may be nonlinearly related across tasks. That is certainly true for verbal stimuli. Even for a simple tone, scale value will depend on whether *S* is set to judge loudness or pitch.

In this view, the subjective magnitudes that are studied in psychophysical judgment may be rather distant derivatives from any basic "sensation." The commonsense view, that there is a direct relation, may well be correct, but it requires justification. Confidence in any measure of sensation would thus seem to rest on a concept of cross-task scale invariance.

Need for such cross-task validation agrees with the emphasis on converging operations by Garner, Hake, and Eriksen (1956), and in an important article by Birnbaum and Veit (in press). Similarly, Weiss and Anderson (1969) commented on scale invariance in psychophysics that "The validity of functional measurement will depend on getting the same scales with different integration tasks [p. 63]." The present and previous reports (Anderson, 1972a) have demonstrated such cross-task scale invariance.

The Psychological Law

The disagreement between magnitude estimation and rating methods is only a surface reflection of a deeper difference. Indeed, since functional measurement explicitly allows for monotone transformation, it could make direct use of magnitude estimation data by transforming out the bias.

The deeper difference appears in the validity criterion. Functional measurement rests on the use of some algebraic model; that provides the scaling frame and the validity criterion. But in the magnitude estimation approach, these integration models have been largely disavowed (Stevens, 1971; see also Anderson, in press-a, Section IV, B, 4).

This difference in methodology comes from a shift in conceptual orientation, away from the psychophysical law and toward the psychological law. The prime concern and emphasis in work with magnitude estimation has been with the psychophysical law, that is, with the relation between subjective magnitude and the physical values of the stimulus. The functional measurement approach is primarily concerned with the psychological law, in particular, with models of stimulus integration. Accordingly, the basic problems are substantive problems of stimulus integration. Measurement in general, and the psychophysical law in particular, are derivative from the psychological law. In this respect (Anderson, 1970b, in press-a), functional measurement is a reaffirmation of the spirit of a long tradition in psychophysical judgment.

REFERENCES

ANDERSON, N. H. Application of an additive model to impression formation. *Science*, 1962, **138**, 817–818.

ANDERSON, N. H. Application of a weighted average model to a psychophysical averaging task. *Psychonomic Science*, 1967, **8**, 227–228.

ANDERSON, N. H. Averaging model applied to the size-weight illusion. *Perception & Psychophysics*, 1970, **8**, 1–4. (a)

ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170. (b)

ANDERSON, N. H. Cross-task validation of functional measurement. *Perception & Psychophysics*, 1972, **12**, 389–395. (a)

ANDERSON, N. H. Looking for configurality in clinical judgment. *Psychological Bulletin*, 1972, **78**, 93–102. (b)

ANDERSON, N. H. Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception*. Vol. 2. New York: Academic Press, in press. (a)

ANDERSON, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*. Vol. 2. San Francisco: Freeman, in press. (b)

ANDERSON, N. H., & JACOBSON, A. Further data on a weighted average model for judgment in a lifted weight task. *Perception & Psychophysics*, 1968, **4**, 81–84.

ANDERSON, N. H., & WEISS, D. J. Test of a multiplying model for estimated area of rectangles. *American Journal of Psychology*, 1971, **84**, 543–548.

BECK, J., & SHAW, W. A. Discrimination of loudness similarity. *Perception & Psychophysics*, 1968, **3**, 105–108.

BIRNBAUM, M. H., & VEIT, C. T. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, in press.

BOGARTZ, R. S., & WACKWITZ, J. H. Polynomial response scaling and functional measurement. *Journal of Mathematical Psychology*, 1971, **8**, 418–443.

CURTIS, D. W., ATTNEAVE, F., & HARRINGTON, T. L. A test of a two-stage model of magnitude estimation. *Perception & Psychophysics*, 1968, **3**, 25–31.

CURTIS, D. W., & FOX, B. E. Direct quantitative judgments of sums and a two-stage model for psychophysical judgments. *Perception & Psychophysics*, 1969, **5**, 89–93.

DAWSON, W. E. Magnitude estimation of apparent sums and differences. *Perception & Psychophysics*, 1971, **9**, 368–374.

ERIKSEN, C. W., & HAKE, H. W. Absolute judgments as a function of stimulus range and number of stimulus and response categories. *Journal of Experimental Psychology*, 1955, **49**, 323–332.

ERIKSEN, C. W., & HAKE, H. W. Anchor effects in absolute judgments. *Journal of Experimental Psychology*, 1957, **53**, 132–138.

FAGOT, R. F., & STEWART, M. Tests of product and additive scaling axioms. *Perception & Psychophysics*, 1969, **5**, 117–123.

FELDMAN, J., & BAIRD, J. C. Magnitude estimation of multidimensional stimuli. *Perception & Psychophysics*, 1971, **10**, 418–422.

GARNER, W. R., & CREELMAN, C. D. Problems and methods of psychological scaling. In H. Helson & W. Bevan (Eds.), *Contemporary approaches to psychology*. Princeton, N.J.: Van Nostrand, 1967.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychological Review*, 1956, **63**, 149–159.

KREUGER, L. E. Apparent combined length of two-line and four-line sets. *Perception & Psychophysics*, 1970, **8**, 210–214.

MARKS, L. E., & CAIN, W. S. Perception of intervals and magnitudes for three prothetic continua. *Journal of Experimental Psychology*, 1972, **94**, 6–17.

POULTON, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, **69**, 1–19.

SHANTEAU, J. C. Component processes in risky decision judgments. Unpublished doctoral dissertation, University of California, San Diego, 1970.

SHANTEAU, J. C. Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology*, 1972, **93**, 63–68.

STEVENS, J. C., & RUBIN, L. L. Psychophysical scales of apparent heaviness and the size-weight illusion. *Perception & Psychophysics*, 1970, **8**, 225–230.

STEVENS, S. S. On the psychophysical law. *Psychological Review*, 1957, **64**, 153–181.

STEVENS, S. S. A metric for the social consensus. *Science*, 1966, **151**, 530–541.

STEVENS, S. S. Issues in psychophysical measurement. *Psychological Review*, 1971, **78**, 426–450.

STEVENS, S. S., & GALANTER, E. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, **54**, 377–411.

WEISS, D. J. Averaging: An empirical validity criterion for magnitude estimation. *Perception & Psychophysics*, 1972, **12**, 385–388.

WEISS, D. J. FUNPOT: A FORTRAN program for finding a polynomial transformation to reduce any sources of variance in a factorial design. *Behavioral Science*, 1973, **18**, 150.

WEISS, D. J., & ANDERSON, N. H. Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, 1969, **82**, 52–63.

WEISS, D. J., & ANDERSON, N. H. Use of rank order data in functional measurement. *Psychological Bulletin*, 1972, **78**, 64–69.

(Received March 21, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 234–243

STIMULUS FAMILIARIZATION AND CHANGES IN DISTRIBUTION OF STIMULUS ENCODINGS[1]

ALLEN R. DOBBS[2]

University of Alberta, Edmonton, Alberta, Canada

Independent groups were given 1 or 10 familiarization trials on integrated (consonant vowel consonant word) or nonintegrated (consonant consonant–consonant) trigrams. The 4 groups ($n = 48$) were then given a stimulus recognition task containing the following 4 types of items: (*a*) the original stimuli, and distractor items having their (*b*) first, (*c*) second, or (*d*) third letters in common with the original stimuli. The probability function for encodings isomorphic with letters from the 3 letter positions was inferred from false identifications. Sixteen *S*s from each group then learned 1 of 3 types of paired-associate (PA) list wherein the stimuli were the familiarized items. Encoding during PA learning was inferred from a stimulated-recall task given after PA learning. Contrary to the 1968 conclusions of Martin, the results indicated that stimulus-only familiarization resulted in selective encoding of nonintegrated stimuli. Although those changes did not affect the *rate* of PA acquisition, familiarization did affect stimulus encoding during PA learning.

The stimulus encoding variability hypothesis advanced by Martin (1968) proposes that the perceptual encoding response to a verbal unit may vary on successive exposures, and that the amount of this variability is negatively correlated with the degree to which the unit is integrated. Highly integrated units such as consonant–vowel–consonant (CVC) words are assumed to be consistently encoded on successive paired-associate (PA) trials, and it is this encoding stability which is postulated to underlie the relatively rapid acquisition of lists having integrated-stimulus items. Lists which contain unintegrated (low meaningfulness, *m*) stimulus units, such as low *m* CCCs, are acquired more slowly, and it is postulated that this decreased learning rate is due to the lack of initial stable encodings for unintegrated units. However, stability of encodings for unintegrated stimuli is assumed to vary across trials. Specifically, it is assumed that there are initial probabilities for the elicitation of the various possible encodings of unintegrated stimuli, that this probability density function (PDF) shifts toward degeneracy for a single encoding early in learning, and then becomes more platykurtic with overlearning. That is, the initial PDF distribution is assumed to shift such that a single selective encoding becomes dominant early in learning, but all encodings become equally probable during overlearning training. Although it is left unspecified as to why or by what mechanism the PDF shifts to degeneracy, it seems clear that Martin considers the cued-recall situation to be a necessary condition. This appears intuitively reasonable, as the functional value of stable selective stimulus encoding during PA learning is obvious, whereas it is difficult to see why individuals would selectively encode at all in a stimulus-only familiarization task. There are also data available which are, at least, not incompatible with this notion. For example, Schulz and Martin (1964) have reported that stimulus familiarization, consisting of aural (spelled) presentation and requiring recall, was found not to have a differential effect on the subsequent rate of PA learning for several levels of stimulus *m*.

On the basis of these data and the Martin (1966) results, Martin (1968) has concluded that increases in stimulus recallability, such as that produced by stimulus-only familiarization, does not affect the acquisition of an

[1] This research was supported by Grant APA 335 from the National Research Council of Canada. Dona Carlson and Marvi Blackmore assisted in conducting the experiment. Bryce Schurr and Susan Schurr assisted in the analyses.

[2] Requests for reprints should be sent to Allen R. Dobbs, Department of Psychology, University of Alberta, Edmonton, T6G 2E1, Alberta, Canada.

association between an encoding(s) and the response term differentially over levels of *m* (or integration). Further, he stated that this

> therefore implies that independent familiarization with a nominal stimulus does not alter the distribution of probability density over $\{r_1 - s_1\}$ [the encodings], at least not in a way that changes the availability of one alternative encoding relative to another for purposes of incorporation into an overt-response-producing association [p. 423].

It is not at all clear that this conclusion is necessitated by the data from the Schulz and Martin (1964) study, for it is possible that the PDF for encodings is altered by stimulus-only familiarization, and this change may carry over to encoding during PA learning in such a way as to systematically alter the availability of possible encodings for incorporation into overt-response-producing associations, *but still not* differentially change the rate of association formation. This interpretation is also consistent with the Schulz and Martin data.

In order to differentiate between these 2 possible interpretations, a more direct test is needed of the effects of stimulus-only familiarization on the PDF for various encodings, and of the type of encoding used during PA learning as a function of the amount of stimulus-only familiarization. Support for Martin's (1968) interpretation would be obtained if the data indicated no change in the PDF for various encodings with increased stimulus-only familiarization, and if there was no difference in stimulus encoding during PA learning as a function of the amount of stimulus-only familiarization.

The present experiment was designed to provide this information. First, independent groups were familiarized for 1 or 10 trials on materials differing widely in the degree to which they are integrated. High *m* CVC words were selected to represent the highly integrated units, and low *m* CCCs were employed as the unintegrated units. The PDF for encodings isomorphic with letters from Positions 1, 2, or 3 of the trigram stimuli was then inferred from the false identifications made during a stimulus recognition task given immediately after familiarization. The effect of stimulus familiarization on the rate of PA learning was then assessed for 3 types of lists. It was assumed that if any shift in the PDF occurred, it would be toward greater selective encoding of the initial position, at least for the unintegrated stimuli. On the basis of this assumption, 3 types of PA lists were constructed and administered to independent groups: For one type of list, selective encoding of the first-position letter would facilitate list acquisition, for another it would lead to interference, and for the third there should be no effect. These conditions were manipulated by varying the relationship of the initial letters of the stimulus and response terms. Finally, a stimulated-recall test was administered in order to assess whether the type of encoding employed during PA learning was a function of stimulus-only familiarization.

Nonrecall familiarization was employed in the present study because it does not impose task demands with regard to stimulus terms which may be in opposition to those of the PA task. That is, familiarization requiring recall would seem to demand that individuals not stimulus (letter) select, whereas in a PA task they tend to do so with nonintegrated stimulus materials (Postman & Greenbloom, 1967). The nonrecall procedure would appear to introduce no biases since the essence of Martin's (1968) conclusions concerning familiarization have to do with stimulus repetition outside the context of a cued-recall situation, and stimulus availability in the sense of recallability is undoubtedly also correlated with the amount of nonrecall familiarization.

Method

Design and subjects. The overall experiment consisted of a $2 \times 2 \times 3$ factorial design. The respective sources of classification were unit integration (high vs. low), number of familiarization trials (1 or 10), and appropriateness of familiarization encoding (same, same re-paired, and different). The appropriateness conditions were formulated primarily for the low-integration (low *m* CCC) stimuli on the basis of 2 assumptions. First, it was assumed that *S*s do selectively encode low *m* CCCs during nonrecall stimulus familiarization, and second, that this encoding would be in accordance with information

theory analyses (Aborn & Rubenstein, 1958; Carson, 1961; Garner, 1962). That is, that there would be a shift in the encoding PDF away from the middle letter, with the predominance of the shift being directed to a selective encoding of the initial letter. On this basis, selective encoding (first letter) was "appropriate" for the same (S) condition, in that paired stimulus and response terms had identical first letters. It was inappropriate for the same re-paired (S_r) condition as the items from the S list were re-paired. In the different (D) condition, all stimuli and responses began with different letters and provided a baseline for evaluating the effects of the S and S_r list constructions.

The *Ss* were 144 University of Alberta undergraduates whose experimental participation was in partial fulfillment of an option for a course requirement, and 48 paid volunteers from an introductory psychology summer course. These latter *Ss* were equally divided among all conditions. Since no differences were detected between the 2 "sets" of *Ss*, they were grouped together in all analyses presented herein. Eight males and 8 females were assigned randomly to each of the 12 groups, with the restriction that each group must have $n - 1$ *Ss* prior to the assignment of the *n*th *S* to any group.

Procedure and materials. The experiment was conducted in 4 phases. Phase 1 consisted of 1 or 10 stimulus familiarization trials on the integrated (CVC) or unintegrated (CCC) materials. A trial consisted of a 2-sec. presentation of each of the 5 stimulus terms. The *Ss* were instructed to read silently each set of 3 letters, as it appeared, but were not asked to memorize the list. During Phase 2, the 4 Phase 1 groups received a stimulus recognition task. The recognition list consisted of 4 kinds of items: (*a*) original, and distractors which had (*b*) first, (*c*) second, or (*d*) third letters in common with the original items. The original items each appeared 3 times, and there were 2 examples of each of the 3 kinds of distractors per original item. These 45 items were presented individually with *S* self-pacing, and ordered randomly with the restriction that the same *type* of item could not occur on 2 successive presentations. The *Ss* were informed of the repetition of original items and instructed to identify the original item by saying *same*, but to respond with *different* when they thought the presented item was not *identical* to an item from Phase 1. *Same* responses to distractor items were taken as evidence for selective encoding on the position of the letter common to an original item.

Each of the 4 groups was divided to form the S, S_r, and D conditions of Phase 3 PA learning. The study–test method was employed with a 2:2-sec. rate of presentation with a 2-sec. intertrial interval.

The Phase 4 stimulated-recall task was identical to the recognition task employed in Phase 2 except that after making each *same-different* discrimination *S* was also instructed to give the first response from the PA list that the item made him think of.

One of the objectives of the present study was to evaluate the stimulus encodings employed during PA learning by means of the Phase 4 stimulated-recall task, and to compare these with [illegible] encoding changes assessed in Phase 2 which might be obtained as a result of differential stimul[illegible]arization. Since the form of encoding PD[illegible] [illegible]ssumed by Martin (1968) to be, in part, a functi[illegible] degree of learning, and since list criteria re[illegible] often widely varying levels of individual-it[illegible] [illegible], an apparatus was constructed to cont[illegible] the variable as carefully as possible. Th[illegible] stimulus materials were presented on IEE Series [illegible] digit display units programmed by means of [illegible]-tape reader. A series of switches allowed fo[illegible] immediate removal and replacement of any [illegible] of the items at any time. When an *S* correctly r[illegible] an item on 2 successive trials, the item wa[illegible] removed from the list and replaced by a pair [illegible] considerably different classes of materials [illegible] interference. The replacement items were [illegible] form–2-digit number pairs. Phase 3 in[illegible] noted the possibility of changing items [illegible] emphasized attending to both types of pairs. [illegible] data indicated that the instructions were successful [illegible] that *Ss* did learn the replacement items, and were able to successfully recall them at the con[illegible] of the experimental session.

Materials. The unintegrated stimulus [illegible] consisted of CCC trigrams having a mean Witmer (1935) *A* value of .33. The integrated sti[illegible] were all 1.00-*a*-value CVC words (Noble, 1961). [illegible] Thorndike and Lorge AA nouns served as respons[illegible]. The distractor stimuli had a mean Witmer *A* [illegible] of .38 and a Noble *a* value of .98 for the low and high stimulus sets, respectively. Since, excluding Y, there are only 5 vowels, the vowels were repeated in the CVC distractors. For comparability, the middle letters of the original-stimulus CCC terms were treated in the same manner. Although this necessitated a confounding in the sense that all types of distractor items had their second letters in common with an original item, no distractor item was allowed to have 2 letters in common with an original item. Further, this confounding should have no differential effect on the assessment of first and third position encoding, nor should any effect be differential for second position between groups, at least within integration levels. The 5 CCCs take 15 consonants leaving only 4 consonants. Only these 4 consonants served as different letters in the distractors. To equate for this, only 4 different consonants were used in the CVC distractors.

RESULTS

Stimulus recognition (Phase 2). The primary objective of Phase 2 was to assess familiarization-induced changes in the PDF for encodings isomorphic with letters from Positions 1, 2, and 3 of the trigram stimuli. The false recognition data shown in Figure 1 are plotted in terms of means of proportions of total recognition errors per *S* on distractor stimuli attributable to identities

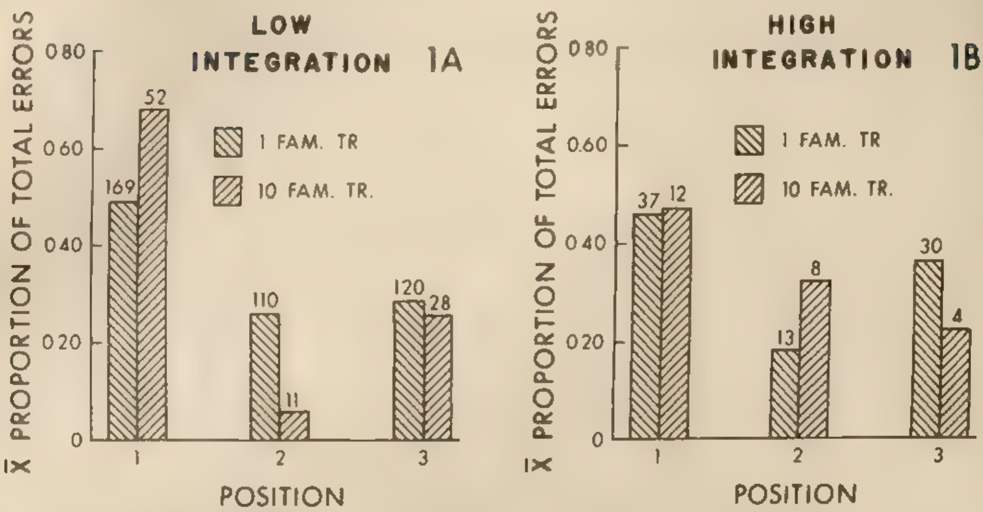


FIGURE 1. Mean proportion of the total number of false recognition errors in Phase 2 attributable to distractors having letters in Positions 1, 2, or 3 identical to original stimuli for low-integration (Figure 1A) and high-integration (Figure 1B) materials. (The numbers represent the frequencies of each type of error.)

in each of the 3 positions. The numbers above the bar graphs in both figures indicate the total number of errors given to each type of distractor item for each of the conditions.

The total number of errors (false identifications) were analyzed using a 2 × 2 × 3 analysis of variance with the factors being integration level (high or low), 1 or 10 familiarization trials, and a within-*S*s variable of first-, second-, or third-letter identity position. The trend shown by the *numbers* in Figure 1 for a greater overall level of number of errors to be made with low-integration stimuli was reliable, F (1, 188) = 86.05, $p < .01$, as was the trend for there to be fewer errors after 10 familiarization trials, F (1, 188) = 76.52, $p < .01$. The interaction of these 2 variables was also reliable, F (1, 188) = 36.68, $p < .01$. The interpretation of the *overall* false identification rates are somewhat ambiguous, as it is unclear whether an individual false identification represents a genuine recognition failure or merely a change in bias for emitting *same* responses as a function of the familiarization conditions. However, any such change in bias cannot differentially affect the within-*S*s *distribution* of errors across the 3 positions, and these are the data of primary concern in evaluating the PDF for the various letter encodings. Similarily, interactions involving the within-*S*s position variable are also unaffected by such a potential change in bias.

The trend for *S*s to differentially distribute their false recognitions across distractor items indicating different types of selective encodings was statistically reliable, F (2, 376) = 26.82, $p < .01$. Most importantly, the distribution of errors depended upon the level of familiarization, as indicated by a significant Familiarization Trials × Position interaction, F (1, 188) = 36.67, $p < .01$. Contrary to Martin's (1968) conclusion, this interaction indicates that there is a change in the PDF for encodings as a function of stimulus-only familiarization. However, the analysis also indicated a statistically reliable Position × Integration Level interaction, F (2, 376) = 9.12, $p < .01$, and because of this interaction the data for high- and low-integration stimuli were further analyzed separately. Further, since the primary interest in evaluating the false identifications was not in the number of errors per se, but rather in terms of, given that an error was made, what kind of encoding did it represent, the remaining analyses are in terms of

the proportions of total errors per S which could be attributed to the 3 position encodings.

As can be seen from the Figure 1A, the proportions of false recognitions for the unintegrated stimuli for Ss given 1 familiarization trial are nearly rectilinearily distributed over the 3 positions. However, 10 familiarization trials resulted in a shift in the distribution of recognition errors away from Position 2 towards Position 1, indicating a trend for selective encoding of the initial letters with increased familiarization. The data from those Ss who made at least one error were analyzed in terms of a 2 × 3 analysis of variance, where number of familiarization trials and a within-Ss position variable were the factors. In this analysis, the overall proportion of errors summed across position for each S is of necessity 1.00. Consequently, no consideration could be made of the between-Ss aspects of the analysis. On the other hand, the position factor and the Position × Number of Familiarization Trials interaction were of direct concern. Since these are both within-Ss comparisons, the lack of between-Ss variance can have no effect on this portion of the analysis. The analysis confirmed the trend for the proportions of errors to be differentially distributed across the 3 positions, F (2, 144) = 47.61, $p < .01$. Further, the reliable interaction, F (2, 144) = 12.97, $p < .01$, indicated a shift in the encoding of unintegrated stimuli with familiarization training. Internal analyses showed that the differences between 1 and 10 familiarization trials for both Positions 1 and 2 were reliable, ts (144) = 3.66 and 3.17, respectively, ps < .01.

Contrarily, proportions of recognition errors for integrated stimuli indicate an initial encoding distribution which is in accordance with previous investigations of positional information of words (e.g., Horowitz, Chilian, & Dunnigan, 1969; Horowitz, White, & Atwood, 1968). That is, the majority of errors were obtained with first- and last-position identity, with identity in the middle position resulting in but few errors. Interestingly, a shift of the error distribution to first, seco[nd, and] third, in order of decreasing percentages [of e]rrors, resulted when 10 familiarization t[rials] were given. This indicates a possible t[ren]d for a shift to serial processing of CV[C] word materials after several familiarizati[on t]rials. However, an analysis identical to [the] one performed on the low-integration d[ata] did not indicate a reliable interaction. Consequently, it must be concluded at [pre]sent that familiarization training does n[ot a]ffect the PDF for the positions of high m [CV]Cs.

It may also be of some interest to ex[a]mine the number of original-item correct [re]cognitions. For the high-integration st[im]uli, the means were 12.77 and 14.46 fo[r] the groups given 1 and 10 familiarization t[ri]als, respectively. These means for the low–1- and low–10-trial groups were 10.15 and 13.52. For all groups SDs ranged [1].47–2.37. The trend for high-integration stimuli to be better recognized was reliable, F (1, 188) = 38.78, $p < .01$. In addition, the trend for the 10-trial task to facilitate recognition of low-integrated stimuli to a greater extent was also statistically reliable, F (1, 188) = 8.70, $p < .01$. However, this interaction was probably the result of a ceiling effect for the high group receiving 10 familiarization trials (38 of 48 Ss showed perfect recognition). An analysis of the total number of *same* responses emitted regardless of whether they were correctly given indicates that it is doubtful that these findings can be attributed to differential bias levels for the 4 groups. A total of 678, 718, 876, and 740 *same* responses were given by the high–1, high–10, low–1, and low–10 groups, respectively. These totals show a trend for the overall number of *same* responses to increase with greater familiarization for the high groups and to decrease with familiarization for the low groups. The interaction was statistically reliable, F (1, 188) = 13.64, $p < .01$. Since the overall trends for the tendency (bias?) to emit *same* responses are opposite to the trends for the *same* responses which indicated correct recognitions, it does not appear that the findings regarding the correct recognition data need be qualified.

In summary, the present data indicate

that nonrecall familiarization trials have several effects: (*a*) False recognitions indicate that after only one familiarization trial the dominance of letter encodings are in the order of first, last, second for integrated (CVC word) items, whereas there is a nearly rectilinear distribution of codings for unintegrated (CCC) stimuli. (*b*) Higher levels of familiarization result in a shift of the probability distribution of letter encodings for unintegrated items to first, last, second in order of decreasing dominance, but have little effect on the probability distribution of letter codings for integrated items. (*c*) Higher levels of familiarization result in more correct identifications of the familiarized items, and this facilitation may be more pronounced regarding unintegrated items.

Paired-associate learning (*Phase 3*). It will be recalled that during PA learning each item was only presented until it was correctly responded to on 2 successive test trials. Consequently, there was a trials-to-criterion score for each item. These item trials-to-criterion scores were averaged for each *S* yielding a mean trials-to-criterion score for each *S*. The data presented in Table 1 and used in all Phase 3 analyses were in terms of group averages of *S* mean trials to reach criterion. These data were analyzed by means of a 2 × 2 × 3 factorial design, where the factors were integration level, familiarization, and appropriateness, respectively. The analysis revealed statistical reliability for the integration level, F (1, 180) = 12.44, $p < .01$, and appropriateness, F (2, 180) = 32.64, $p < .01$, variables, as well as for the Integration Level × Appropriateness interaction, F (2, 180) = 4.86, $p < .01$. None of the remaining tests approached significance (largest $F = 1.004$). Since the dropout procedure was employed, a similar analysis was conducted which excluded the contribution of the last-learned item. This analysis revealed the same pattern of results except that the potential Integration Level × Familiarization × Appropriateness interaction was more exaggerated though not statistically reliable, F (2, 180) = 2.09. Although there is a trend for both the

TABLE 1

Phase 3 Paired-Associate Performance (Group Averages of *S* Mean Trials to Criterion) as a Function of the Relationship between the Stimulus and Response Term First Letters, Amount of Prior Stimulus Familiarization, and Stimulus Integration

Paired-associate list	Low integration		High integration	
	Amount of familiarization		Amount of familiarization	
	1	10	1	10
Same	2.44	2.65	[illegible]	[illegible]
Same re-paired	5.76	5.29	[illegible]	[illegible]
Different	4.70	4.95	[illegible]	[illegible]

high- and low-integration-S, conditions to be affected by familiarization, neither of these trends was statistically reliable in either analyses including all 5 items or ones excluding the last-learned item. Consequently, it was concluded that the integration level and appropriateness conditions affected learning, but that familiarization had no effect on the rate of PA learning. These findings are in accordance with the results reported by Schulz and Martin (1964) and Martin (1966). Consequently, the data from Phases 2 and 3 taken together indicate that stimulus-only familiarization of unintegrated units does affect the PDF for letter encodings, and that this change in encoding dominance does not necessarily lead to an alteration in the *rate* of PA list acquisition.

Stimulated recall (*Phase 4*). The primary objective in analyzing the stimulated-recall data was to assess false recognitions of the distractor stimuli which were followed by a PA-list response associated with the critical letter of the distractor during Phase 3 learning. Within the context of Martin's (1968) stimulus encoding variability hypothesis, the critical data are any evidence of differences in the distribution of such responses across the various types of distractors as a function of familiarization. Should such differences be realized, they would imply that stimulus-only familiarization does result in a change in the PDF for the encodings (as the Phase 2

TABLE 2

Mean Number of False Identifications in Phase 4 by the Low-Integration Groups which Were Followed by Responses Previously Associated with an Original Item Having a Letter in Common with the Distractor

Critical letter position	1 familiarization trial			10 familiarization trials		
	Same	Same repaired	Different	Same	Same repaired	Different
1	6.44	1.19	1.62	1.69	.56	.75
2	.06	.19	.19	.12	.06	.06
3	.19	.56	.25	.06	.25	.38

data indicated), and familiarization does in fact do so in a way which alters the availability of alternative encodings for purposes of incorporation into overt-response-producing associations. A finding of this type would indicate an effect of familiarization on PA learning (presumably the stimulus encoding phase) even though the *overall* rate of list acquisition was unaffected.

Those *S*s receiving the familiarized integrated items (CVC words) as stimulus terms during the PA task made very few recognition errors to the distractor stimuli which were followed by a "correct response." Only a total of 20 such errors were made by 92 *S*s, and these were not differentially distributed across conditions. It appears then, that stimulus-only familiarization of the integrated stimuli had little effect on the PDF for the various letter encodings (Phase 2 data), and there was no effect of familiarization on encoding during PA learning. This finding is not surprising as it would be anticipated that high *m* CVC words are well integrated and this integration would be little affected by familiarization training.

Contrarily, those *S*s receiving the unintegrated stimuli (low *m* CCCs) made a total of 234 false identifications which were followed by a PA-list response previously associated with an original stimulus item having a letter in common with the distractor. Further, these responses were differentially distributed with respect to the 3 letter positions and the appropriateness and familiarization conditions. Table 2 shows the means for these comparisons. As is apparent from Table 2, there were very few false identifications followed by correct responses to any distractor items, except those having their first letters in common with an original stimulus. Because of the exceedingly small number of identification errors representing encoding on Positions 2 or 3, only the data from the Position 1 items were analyzed. A 2 × 3 (Familiarization Level × Appropriateness) analysis of variance indicated that prior stimulus-only familiarization affected the number of Position 1 false identifications followed by a correct response which were made after PA learning, F (1, 90) = 41.03, $p < .01$. As can be seen from Table 2, each of the 10-familiarization-trial groups made fewer such errors than did their respective 1-familiarization-trial groups. Since very few Position 2 and 3 items were falsely identified, this finding indicates that the 10-familiarization-trial groups knew more about the stimulus items than did the 1-trial groups, whose knowledge was more restricted to the initial letters. Since all groups had reached similar levels of learning during the PA task, this finding may reflect a more platykurtic PDF for the groups receiving a greater degree of stimulus-only familiarization.

The type of PA list also influenced the number of Position 1 false identifications which were followed by a correct response, F (2, 90) = 38.91, $p < .01$, and the amount of familiarization differentially affected *S*s receiving different types of PA lists, F (2, 90) = 16.85, $p < .01$. These findings further indicate that the amount of stimulus-only familiarization did affect the encoding during PA learning.

A consideration of the responses to the original stimuli provides for an assessment of recall as a function of recognition performance. The number of correct identifications (CIs) regardless of whether they were accompanied by the correct PA-response terms was compared with the number of CIs that were followed by elicitations of correct response terms from

the PA list (CI–CR). This within-*Ss* factor was combined with 3 between-*Ss* comparisons of the integration level, familiarization, and appropriateness variables. Overall, the design was a 2 × 2 × 3 factorial arrangement. Table 3 shows these conditions and corresponding performance levels. Of the between-*Ss* sources of variance, the integration level and appropriateness factors yielded statistically reliable differences, $F\ (1, 180) = 8.02$ and $F\ (2, 180) = 9.03$, respectively, $ps < .01$. The familiarity variable and Integration Level × Appropriateness interaction approached reliability, $F\ (1, 180) = 3.82$ and 3.28, respectively, $ps < .10$. The reliable within-*Ss* variable of CI vs. CI–CR, $F\ (1, 180) = 120.00$, $p < .01$, and CI vs. CI–CR × Appropriateness interaction, $F\ (2, 180) = 26.96$, $p < .01$, indicate that given stimulus recognition, *Ss* still had trouble recalling the responses, and the recall difficulty depended upon the PA condition. No other factors or interactions produced *F*s greater than unity.

TABLE 3

MEAN NUMBER OF CORRECT IDENTIFICATIONS (CI) AND MEAN NUMBER OF CIs FOLLOWED BY A CORRECT RESPONSE (CI–CR) IN PHASE 4 AS A FUNCTION OF INTEGRATION LEVEL, NUMBER OF PRIOR FAMILIARIZATION TRIALS, AND PAIRED-ASSOCIATE CONDITION

Familiarization trials and paired-associate condition	Low integration		High integration	
	CI	CI–CR	CI	CI–CR
1				
Same	12.69	12.31	12.88	12.00
Same re-paired	13.62	8.12	11.19	10.62
Different	13.31	10.00	11.69	12.00
10				
Same	13.50	13.50	13.94	13.38
Same re-paired	13.44	9.31	12.81	9.69
Different	13.94	10.32	12.31	12.81

DISCUSSION

Taking all phases into account, the following conclusions emerge: (*a*) The error data indicate that nonrecall stimulus familiarization of unintegrated stimuli does lead to changes in the selective encoding of the familiarized items from rectilinearity to an ordering consistent with information analyses of words (first, last, second, in order of decreasing dominance). (*b*) There seems to be little evidence of a shift in encoding integrated stimuli as a result of familiarization, although any shift would appear to be away from first, last, middle, and in the direction of serial processing. (*c*) Familiarization results in an overall increase in the number of correct recognitions and a decrease in the number of false recognitions for both integrated and unintegrated stimuli. (*d*) Although familiarization induced encoding changes do occur, they appear to have little effect on the *rate* of PA learning. (*e*) A stimulated-recall task given after PA learning indicated that although 1 and 10 trials of familiarization lead to the same rate of obtaining correct responses in PA, the type of encoding may be quite different for unintegrated stimuli. The implications of the present findings for the stimulus encoding variability hypothesis will be considered below.

The data from Phase 2 indicated that for unintegrated stimuli a shift in the encoding PDF occurs during nonrecall familiarization, even though there would appear to be no "reasons" for this to happen. In fact, given the simple instructions to read the items, it is unlikely that *Ss* did not suspect a second task such as recalling the items they had read. Since this type of *S* hypothesis should have led to rectilinear PDFs, the reliable shift to first-position dominant occurred in spite of this potential problem. The finding of an altered PDF as a consequence of stimulus-only familiarization is clearly contrary to conclusions reached by Martin (1968). Although it is important to note that he was explicitly talking about familiarization procedures which altered stimulus availability in the sense of recallability, there can be little doubt that in the present study the familiarization variable would have produced differences in stimulus recall had such a measure been taken. However, in order to assess this assumption directly, 4 groups of 16 *Ss* each were administered the familiarization task, and then asked to recall the items. The obtained recall means were 1.50, 3.50, 4.31, and 5.00, for the low-1-trial, low–10-trial, high–1-trial, and high–10-trial groups, respectively. As can be seen, the familiarization task employed herein was effective in altering stimulus availability in the sense of recallability. Furthermore, the means for the recall task directly parallel those obtained with the recognition procedure. Thus, although it appears counterintuitive,

the present findings suggest that mere repetition of nonintegrated verbal items results in a shift in the PDF for letter encodings toward first-position dominance.

The obtained focusing on the first-position-letter encoding for the low–10-trial groups was expected to facilitate PA learning for the S groups as compared to the respective low–1-trial group. That the facilitation was not obtained for the low-integrated S comparison was probably due to a ceiling effect as further evidenced by the lack of difference between any of the low- and high-integration S groups on the PA learning task. However, equivalent rates of PA acquisition after varying levels of stimulus familiarization should not be taken to indicate that familiarization does not affect stimulus processing during PA learning, as has previously been concluded by Martin (1968). The Phase 4 stimulated-recall data from the present experiment indicated that the low–1-trial S group poorly discriminated between original stimuli and distractors which had first letters in common with these stimuli. Contrarily, the analogous 10-trial group rarely made this type of error, but rather performed very much like the high-integrated S groups in that these *S*s were not only able to identify the original stimuli, but also most often correctly identified the distractors and gave a response which would have been correct on the basis of the first letter. The percentages for these contingencies were 23.12, 64.38, 50.00, and 43.75 for the low–1- and low–10-, and high–1- and high–10-S groups, respectively. As can be seen, the low–1-trial group is substantially different from the remaining 3 groups. This indicates the low–10-trial S group knew the relationship between the stimuli and responses (first letters identical), and also knew something about the entire stimulus items. Contrarily, the low–1-trial S groups apparently knew little about the stimulus items except the first letters. In terms of the hypothesis advanced by Martin (1968), these data imply that for low-integrated items the PDF degenerated toward a single encoding with increased familiarization, and may have become platykurtic more rapidly during PA learning than is the case with little or no familiarization. Hence, it is suggested that the effect of nonrecall stimulus familiarization may be either to speed up the encoding phase during PA learning, or to allow a considerable portion of this phase to occur prior (and transfer) to the PA task. Of these, the latter seems more appealing at present.

Although greater familiarization of low-integrated items was expected to lead to increased interference within the S, cond[illegible] in the opposite direction [illegible] lack of increase in interf[illegible] occurred as a result of a domin[illegible] encoding leading to a more ra[illegible] of the interference and a compen[illegible] in encoding. However, it seems [illegible] this type of shift occurs. Williams[illegible] wood (1970) have reported that [illegible] spontaneously shift encodings away fr[illegible] position to avoid interference, at least [illegible] 2-list situation. However, the trend fo[illegible] creased interference in the low-integrated condition with increased familiarization wo[illegible] be consistent with the hypothesis that wi[illegible] increased familiarization, the encoding PDF may become platykurtic more rapidly during a PA task. This would also tend to reduce interference due to re-paired stimulus and response first letters. The trend for increased familiarization of high-integrated stimuli to lead to increased interference is a puzzling one.

Throughout this report, the emphasis has been on the distribution of encodings isomorphic with the letter components. This is, of course, an exceedingly limited view. Other plausible encodings would include 2-letter combinations as well as an encoding of a unitized version of the item. Within this context, one could consider all of the items presented on the stimulus recognition trials to constitute tests of encodings. The 3 types of distractor items would be assumed to test for individual letter selection, or at least letter dominance, and the original items as tests of stimulus integration encodings. Unfortunately, the latter is confounded, as a correct recognition of an original item could be based on single-letter encodings, 2-letter combination encodings, or encodings based on all 3 letters. Nevertheless, this confounding can be made somewhat less ambiguous by the following analysis. If familiarization results in only a shift to single-letter encodings based on the first position for low-integration materials, then one might expect not only an overall increase in the number of correct recognitions of original stimuli (as was obtained) but also an increase in the number of false recognitions to distractors having first-position letters identical to the original stimuli (the opposite was obtained). Familiarization resulted in an increase in the *proportion* of false identifications of Position 1 distractors but a decrease in the overall number of correct identifications.

This finding would be anticipated if the PDF for the various encodings was assumed to shift from some initial distribution to a dominance of a single encoding, and finally to an integrated version of the nominal stimulus as learning processed during familiarization training. If the rate of these encoding changes was less than uniform for all items, the higher level of familiarization would be expected to result in more correct identifications and fewer errors; but given an error, they would be primarily those indicating first-position encodings, as was obtained. Although the Phase 2 data indicated changes in the distributions of encodings, the rate of PA acquisition obtained in Phase 3 was not a function of the amount of prior stimulus-only familiarization. Nevertheless, the Phase 4 recognition data clearly indicate that the shift in the PDF for the various encodings resulting from stimulus-only familiarization did transfer to the PA task, and in such a way as to alter the availability with which the alternative encodings could enter into overt-response-producing associations.

REFERENCES

Aborn, M., & Rubenstein, H. Perception of contextually dependent word probabilities. *American Journal of Psychology*, 1958, **71**, 420–422.

Carson, D. H. Letter constraints within words in printed English. *Kybernetik*, 1961, **1**, 46–54.

Garner, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.

Horowitz, L. M., Chilian, P. C., & Dunnigan, K. P. Word fragments and their redintegrative powers. *Journal of Experimental Psychology*, 1969, **80**, 392–394.

Horowitz, L. M., White, M. A., & Atwood, D. W. Word fragments as aids to recall: The organization of a word. *Journal of Experimental Psychology*, 1968, **76**, 219–266.

Martin, E. Stimulus pronunciability in aural paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 18–22.

Martin, E. Relation between stimulus recognition and paired-associate learning. *Journal of Experimental Psychology*, 1967, **74**, 500–505.

Martin, E. Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review*, 1968, **75**, 421–441.

Noble, C. E. Measurements of association value (a), rated associations (a'), and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. *Psychological Reports*, 1961, **8**, 487–521.

Postman, L., & Greenbloom, R. Conditions of cue selection in the acquisition of paired-associate lists. *Journal of Experimental Psychology*, 1967, **73**, 91–100.

Schulz, R. W., & Martin, E. Aural paired-associate learning: Stimulus familiarization, response familiarization, and pronunciability. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 139–145.

William, R. F., & Underwood, B. J. Encoding variability: Tests of the Martin hypothesis. *Journal of Experimental Psychology*, 1970, **86**, 317–324.

Witmer, L. R. The association value of three-place consonant syllables. *Journal of Genetic Psychology*, 1935, **47**, 337–360.

(Received April 12, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 244–249

PERIPHERAL VISION LOCATION AND KINDS OF COMPLEX PROCESSING

DAVID C. EDWARDS[1] AND PAULA A. GOOLKASIAN

Iowa State University

Processing capacity in the visual periphery was tested with 4 kinds of tasks: (*a*) detection of light onset, (*b*) recognition of the position of a U-shaped figure, (*c*) identification of a letter, and (*d*) categorization of 3-letter words. These tasks were presented at 2 luminances, 3 sizes, and 3 stimulus durations at distances from the fovea (10°, 15°, 25°, and 58°) at which retinal structures differ. In the near periphery, 10°, performance under optimal conditions was good for all tasks. At 15° information concerning detection, recognition, and identification was handled with a high degree of accuracy; however, the classification task was performed at a level just above chance. The processing capacity in the far periphery, 25°, was good only for detection judgments. At 58°, extreme periphery, the detection task was the only task performed above chance. Of the state variables, size had an important effect at all locations, but brightness and stimulus duration had minimal effects. Response time measures gave supporting evidence for the accuracy of response data.

Conceptions of peripheral vision have come from studies of acuity (e.g., Kerr, 1971; Low, 1943, 1947; Mandelbaum & Sloan, 1947; Wertheim, 1894), visual search (e.g., Mackworth & Morandi, 1967; Williams, 1966; Yarbus, 1967), and descriptions of the kinds of processing limits upon peripherally received stimulation (Engel, 1971; Leibowitz, Johnson, & Isabelle, 1972; Mackworth, 1965; Menzer & Thurmond, 1970; Newsome, 1972; Payne, 1967; Sanders, 1970).

The salient fact of peripheral vision has been the marked effect on measures of acuity. The dominance of the acuity effect has inhibited study of other kinds of capabilities and limits of peripheral vision. There have been only a few writers who have accepted the acuity facts and continued to ask questions of the role of peripheral vision. According to Gibson (1966), the periphery functions to guide the fovea: "Interesting structures in the array, and interesting bits of structure, particularly motion, draw the fovea toward them, [p. 360]." In this context, stimulation reaching the periphery is presumed to be processed as a locus of concentration of information within a visual display, but the information itself is not fully analyzed. Mackworth and Morandi (1967) presented data that supplement Gibson's notion. They found a greater number of fixations in the regions of the pictures that contain unpredictable contours or unusual information. Thus the amount of information in each locus determined the adjustments of the eye. Peripheral vision in their account processes predictable and redundant information. They further relied on Hubel and Wiesel (1965) for physiological evidence to suggest that there is a rapid processing in the periphery of straight lines and slightly curved boundaries. Mackworth and Morandi summarized the nature of the visual mechanism by writing " . . . the peripheral retina therefore quickly screens off the predictable features and leaves the fovea to process the unpredictable and unusual stimuli [1967, p. 551]."

These notions suggests a processing dichotomy in the visual system: fovea vs. periphery. Yet attempts at empirically mapping the processing ability of the periphery do not support that notion. The classical acuity functions based on Landolt ring discrimination have shown performance to decrease rapidly from the fovea to 5° with a slower decline in performance over the remaining area (e.g., Wertheim, 1894). Sanders (1963) offered specifications for

[1] Requests for reprints should be sent to David C. Edwards, Department of Psychology, Old Botany Hall, Iowa State University, Ames, Iowa 50010.

functionally different areas within the visual field based on a light discrimination task. Following a performance curve[2] that deviated from a linear function of visual angle at 2 points, at the region of 20°–25° from the fovea and at 80°, he formulated a tripartite system of the visual field. In the stationary field, particularly the area extending from the fovea to 20°–25°, incoming signals are grouped and processed by foveal and peripheral vision. The eye field uses eye movements to handle information in the area out to 80°. The eye field processes in some form between grouping and successive handling of signals. The head field includes the area beyond 80° and treats incoming signals successively using both head and eye movements. The importance of Sanders' distinction lies not in its delineation of specific peripheral fields but rather in its theoretical movement away from describing the visual system as strictly peripheral vs. foveal and in its investigation of processing types within the periphery itself. Polyak's (1941) careful summary of the change in retinal structure through the periphery provides another consideration for assuming a continuous or at least multiple-staged retina.

The studies described below are an exploratory attempt at comparing different processing tasks at several locations in the periphery. Variables of luminance, size, and duration make up a matrix of state conditions to attempt to locate effective variables for peripheral processing. There are 4 studies and each compares 4 tasks at 2 luminances, 3 stimulus sizes, and 3 exposure durations. Since the experiments differ only in peripheral location and *S*s, they are reported together. The results can be examined for evidence of the maximum processing capacity at each location and also of the extent to which state variables affect performance.

Method

A within-*S*s factorial design was applied to separate groups of 10 *S*s at each of 4 peripheral angles: 10° (the foveal edge of the near periphery), 15° (the foveal edge of the middle periphery), 25° (the foveal edge of the far periphery), and 58° (the foveal edge of the extreme periphery). It was assumed that these locations would best represent 4 peripheral regions distinguished neurologically by Polyak (1941). The 4 stimulus tasks used were presumed to vary in processing complexity. These were (*a*) detection of a lighted line segment, (*b*) recognition of the position of a U-shaped figure, (*c*) identification of a letter, and (*d*) categorization of 3-letter words. Three exposure durations, 3 stimulus sizes (distances), and 2 stimulus luminances completed the design. The tasks were given on successive days in the above order to all *S*s.

Apparatus. The visual tasks were presented by alphanumeric electroluminescent displays in which elements were separately programmed. The light detection task consisted of one lighted segment of the display measuring 4.75 × 34.90 mm. The stimuli of the recognition task were U-shaped figures measuring 25.40 × 34.90 mm. and appearing in one of 4 positions: upright, inverted, opening to the right, and opening to the left. The stimuli of the identification task were the letters H, L, O, and X, 50.8 mm. high and 34.9 mm. wide. The categorizing task used 3 adjacent displays to present a 3-letter word, 50.8 mm. high and 162.4 mm. wide. The words belonged to one of 4 categories: animals, part of body, proper names, and utensils.

The 3 exposure durations were 200, 600, and 1,000 msec. Stimulus size was varied by placing the display at 1.83, 2.74, or 3.66 m. from *S*'s eye. The luminance of the lighted elements was 3.9 cd/m² or .043 cd/m² upon a black background of .0017 cd/m², measured by Spectra brightness spot meter, model UB 1½°.

The fixation point was an illuminated (.48 cd/m²) ball, 3.81 mm. in diameter, 150 mm. above the center of the display, displaced to one side so as to make the experimental angles of 10°, 15°, 25°, or 58°. The *S*s were seated in an adjustable chair enabling easy control over the distance between *S*'s eye, the fixation point, and the displays. The chair also provided a brace against which *S* could restrict head movements. A black eye patch was always placed on the right eye, and the displays were presented to the temporal visual field of the left eye.

The *S*, holding a response box containing 4 buttons, indicated the response by depressing the appropriate button. This also provided a response time measure. An association between the stimulus alternatives and the position of the buttons was formed during the practice trials for each task.

Electrooculograms were recorded from Beckman Biopotential electrodes 4 mm. in diameter affixed to the skin about 5 mm. to the temporal side of each eye. The Beckman Dynograph recording provided a record assuring that *S* made no saccades greater than 5° at the time of each stimulus presentation.

Subjects. The *S*s were 40 men and women student volunteers who were given credit towards their general psychology course grade. They all had

[2] These measures were taken on a binocular visual field whereas most of the Landolt ring studies were performed using a monocular field.

TABLE 1

Mean Number of Correct Responses

Conditions	Visual angles			
	10°	15°	25°	58°
Task				
1	7.99	7.33	8.00	7.98
2	6.17	4.19	3.16	2.07
3	6.31	4.54	4.27	2.39
4	3.32	2.15	1.93	1.63
Stimulus size				
Large	7.05	5.89	5.07	3.73
Medium	5.76	4.22	4.23	3.37
Small	5.03	3.55	3.72	3.45
Stimulus brightness				
.043 cd/m²	5.63	4.31	4.16	3.46
3.9 cd/m²	6.27	4.80	4.52	3.58
Stimulus duration				
.2 sec.	5.82	4.49	4.23	3.46
.6 sec.	5.98	4.50	4.35	3.59
1.0 sec.	6.05	4.67	4.35	3.51

normal vision in their left eye without glasses or contact lenses. The *S*s were randomly assigned to each of the 4 peripheral locations. Apparatus design dictated that the 5 *S*s run on any 1 day belong to the same group.

Procedure. Each *S* was given one of the 4 tasks in daily test sessions of 1 hr. Within each 144-trial test session, the stimuli were presented a total of 8 trials at each factorial combination of size, brightness, and exposure duration. The presentation order was counterbalanced across *S*s for the size of the stimulus (half of the *S*s received the sequence at distances of 1.83, 2.74, and 3.66 m. and the others in the reverse order), and the presentation order of each brightness and exposure duration was random within each size level.

The *S*s were visually familiarized with the stimulus forms and given at least 15 practice trials with feedback. However, in the word categorization task, they were given verbal examples of possible members of the 4 categories and then given 15 practice trials with feedback.

The *S*s were instructed to fixate their eye on the illuminated ball, focus their attention to the periphery, and press the appropriate button as soon as they detected the stimulus. Any *S* found by the electrooculograph to have moved his eye fixation during stimulus exposure was removed from the experiment. Feedback as to the correctness of the response was not given during the experimental session.

Results

A separate analysis was made for each angle and for each dependent measure: number of correct responses and response time. A factorial design was used having *S*s as the only random factor (3 levels of size × 2 levels of brightness × 3 exposure durations × 4 tasks × 10 *S*s). The magnitude of the experimental effects was expressed as ω^2, indicating the ratio of that variance component to the sum of the variance components of that factor and its error term (see Hays, 1963; and Vaughan & Corballis, 1969). The main effects of all 4 experiments (angles) are presented in Table 1 (number correct) and Table 2 (response time).

10° visual angle. Performance in the near periphery at 10° was high for the 4 tasks. For the detection task (Task 1), the *S*s made 100% correct judgment. They were 77% correct on the recognition task (Task 2), 79% correct on the identification task (Task 3), and 41% correct on the classification task (Task 4). The strongest of the factors was due to tasks, accounting for 48% of the variance, $F\ (3, 27) = 82.54$, $p < .005$, $\omega^2 = .48$. Performance levels among the tasks were in the expected direction; however, Tasks 2 and 3 did not differ significantly, $F > 1$. Tasks differences interacted significantly with size, $F\ (6, 54) = 14.63$, $p < .005$, $\omega^2 = .04$, and brightness, $F\ (3, 27) = 19.66$, $p < .005$, $\omega^2 = .01$, but these interactions are probably due to ceiling effects in Task 1 and possible floor effects on the measure in Task 4.

Of the 3 presentation variables, stimulus size (distance) appeared to have the greatest effect, $F\ (2, 18) = 70.17$, $p < .005$, $\omega^2 = .12$. Brightness was not a strong factor, $F\ (1, 9) = 62.28$, $p < .005$, $\omega^2 = .02$, but it did interact with stimulus size, $F\ (2, 18) = 16.63$, $p < .005$, $\omega^2 = .01$, so that brightness had a greater effect at smaller display sizes. Finally, although performance increased with longer exposure durations, $F\ (2, 18) = 4.45$, $p < .05$, $\omega^2 = 0.00$, it accounted for a negligible amount of variance.

The analysis of response times supported the findings with number of correct responses. The mean response time increased with the complexity, judged by *E*, of the processing task: .63, 1.29, 1.63, and 2.36

sec., F (3, 27) = 40.58, $p < .005$, $\omega^2 = .59$. Again there was no significant difference between Tasks 2 and 3, $F > 1$. Exposure duration was positively related to response time, F (2, 18) = 16.77, $p < .005$, $\omega^2 = .01$. Levels of brightness, F (1, 9) = 9.23, $p < .025$, $\omega^2 = .01$, and size, F (2, 18) = 25.22, $p < .005$, $\omega^2 = .04$, were inversely related to response time. These latter findings were consistent across the first 3 tasks. The fact that response time did not increase with a decrease in size or a decrease in intensity of the display in the classification task indicates that the *S*s had perhaps lost confidence in their ability to perform the task. That explanation perhaps also accounts for the significant Task × Exposure interaction, F (6, 54) = 7.68, $p < .005$, $\omega^2 = 0.00$, Task × Brightness interaction, F (3, 27) = 8.50, $p < .005$, $\omega^2 = 0.00$, and Task × Size interaction, F (6, 54) = 8.51, $p < .005$, $\omega^2 = .03$.

15° visual angle. In the middle periphery, 15°, accuracy of processing was high for the first 3 tasks, 91%, 52%, and 57%, but low for the classification task, 27%. Some decrement in performance at 15° may be attributed to the blind spot. The strongest factor was task differences. Performance was inversely related to task complexity, F (3, 27) = 79.79, $p < .005$, $\omega^2 = .48$, although Tasks 2 and 3 did not differ, $F > 1$. Size, F (2, 18) = 88.47, $p < .005$, $\omega^2 = .14$, accounted for a greater portion of the variance than did brightness, F (1, 9) = 23.18, $p < .005$, $\omega^2 = .01$, and exposure duration was not a significant effect, F (2, 18) = 1.69. Size and task again interacted, F (6, 54) = 10.41, $p < .005$, $\omega^2 = .06$, as did task and brightness, F (3, 27) = 3.79, $p < .025$, $\omega^2 = 0.00$. Both of those interactions are probably an artifact of a ceiling effect on the response measure in Task 1 and a floor effect in Tasks 3 and 4. The response time data were consistent with those from 10°. Mean response time increased with task complexity: .76, 1.56, 1.76, and 2.37 sec., F (3, 27) = 23.76, $p < .005$, $\omega^2 = .39$. Response time was positively related to exposure duration, F (2, 18) = 13.85 $p < .005$, $\omega^2 = .01$, and was inversely related to size, F (2, 18) = 5.84, $p < .025$, $\omega^2 = .01$. A lack of consistency in effects was noted in the response times for Task 4 as compared to those from the other tasks. Again, this was probably due to *S*'s confidence that he could perform the task. That also explains the Size × Task interaction, F (6, 54) = 7.21, $p < .005$, $\omega^2 = .03$.

TABLE 2

MEAN RESPONSE TIME

Conditions	Visual angles			
	10°	15°	25°	58°
Task				
1	.63	.76	.68	.61
2	1.29	1.56	1.59	1.81
3	1.63	1.76	1.75	1.96
4	2.36	2.37	2.33	2.25
Stimulus size				
Large	1.29	1.48	1.52	1.69
Medium	1.45	1.67	1.61	1.67
Small	1.70	1.69	1.63	1.61
Stimulus brightness				
.043 cd/m²	1.55	1.58	1.57	1.63
3.9 cd/m²	1.41	1.65	1.60	1.69
Stimulus duration				
.2 sec.	1.38	1.49	1.44	1.51
.6 sec.	1.46	1.61	1.60	1.68
1.0 sec.	1.60	1.75	1.72	1.78

25° visual angle. In the far periphery, 25°, the accuracy of performance levels for the 4 tasks was 91%, 40%, 53%, and 24%. The strongest factor was again task differences, F (3, 27) = 191.52, $p < .005$, $\omega^2 = .98$, although Tasks 2 and 3 did not differ, $F > 1$. There was again a size effect, F (2, 18) = 60.6, $p < .005$, $\omega^2 = .06$, a brightness effect, F (1, 9) = 17.66, $p < .005$, $\omega^2 = .01$, a Task × Size interaction, F (6, 54) = 19.26, $p < .005$, $\omega^2 = .05$, and a Task × Brightness interaction, F (3, 27) = 11.23, $p < .005$, $\omega^2 = .01$, but each of these effects was relatively weak.

The response time data at 25° were consistent with those of the other peripheral regions. Mean response time increased with task complexity: .67, 1.59, 1.74, and 2.33 sec., F (3, 27) = 61.47, $p < .005$, $\omega^2 = .64$. Again response time was found to

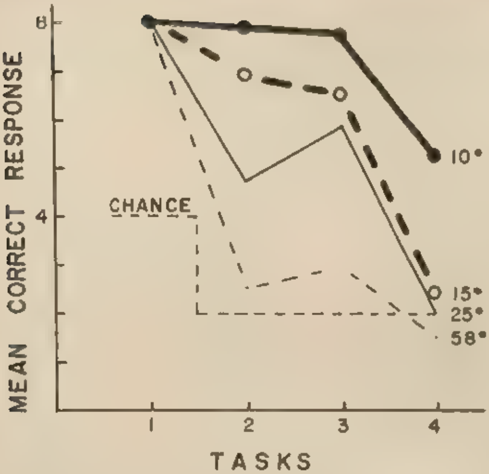


FIGURE 1. Selected mean performance to reflect the optimal (best) performance on each task at each angle; the points are comparable only in that sense.

be positively related to exposure duration, $F\ (2, 18) = 30.07$, $p < .005$, $\omega^2 = .02$, and negatively related to the size of the stimulus, $F\ (2, 18) = 4.73$, $p < .025$, $\omega^2 = 0.00$. Variability was again evident within response time in the fourth task, and as before this is likely accounted for by chance performance levels of the *S*s under all experimental conditions. As in the more nearly foveal groups it explains the interactions of task and size, $F\ (6, 54) = 4.95$, $p < .005$, $\omega^2 = .01$, task and exposure duration, $F\ (6, 54) = 6.50$, $p < .005$, $\omega^2 = 0.00$, and task and brightness, $F\ (3, 27) = 5.99$, $p < .005$, $\omega^2 = .01$.

58° visual angle. Performance in the extreme periphery, 58°, even under optimal conditions was low except for simple light detection in Task 1. The best performances were 100%, 26%, 31%, and 19%. Task, $F\ (3, 27) = 1{,}222.09$, $p < .005$, $\omega^2 = .99$, and size, $F\ (2, 18) = 7.60$, $p < .005$, $\omega^2 = 0.00$, were the only significant main effects. The Brightness × Exposure Duration, $F\ (2, 18) = 5.88$, $p < .025$, $\omega^2 = 0.00$, and Size × Task, $F\ (6, 54) = 3.66$, $p < .005$, $\omega^2 = 0.00$, interactions were also significant but with very small factor strength. These interactions are probably influenced by the Task 1 ceiling effect and Task 4 floor effect on the measure. Response time showed the usual increase with task complexity: .61, 1.81, 1.96, 2.25 sec., $F\ (3, 27) = 49.40$, $p < .005$, $\omega^2 = .53$. Response time also increased with exposure duration, $F\ (2, 18) = 21.11$, $p < .005$, $\omega^2 = .02$, and there were minor effects of size, $F\ (2, 18) = 4.64$, $p < .025$, $\omega^2 = 0.00$, Size × Brightness interaction, $F\ (2, 18) = 6.62$, $p < .01$, $\omega^2 = 0.00$, Task × Duration interaction, $F\ (6, 54) = 4.77$, $p < .005$, $\omega^2 = 0.00$, Task × Brightness interaction, $F\ (3, 27) = 6.44$, $p < .005$, $\omega^2 = 0.00$, and Size × Task interaction, $F\ (6, 54) = 2.92$, $p < .025$, $\omega^2 = .01$.

DISCUSSION

The ability of the periphery to process information depends upon the region stimulated, the complexity of the task, and the size of the stimulus. Figure 1 is a mapping of the processing accuracy at each of the 4 peripheral regions under the most optimal values of the significant factors in the experiments. In the near periphery all 4 types of information were processed, given a large, bright, and long-duration stimulus. At the middle periphery information concerning detection, recognition, and identification were handled at a high degree of accuracy, but the more complex task, classification of words, was performed only at about chance. The processing capacity in the far periphery was good only for detection judgments. It dealt with information about recognition and identification but only at a limited degree of accuracy. At 58° in the extreme periphery the only task that was performed was light detection.

Across the periphery the factor which had the strongest effect on peripheral vision, in addition to task complexity, was size of the stimulus. The effect of size is in accord with the known facts of neural anatomy, especially the distribution of rods in the periphery and their manner of connection to the bipolar and ganglionic cells. Unlike the almost one-to-one connection with the foveal cones, the bipolar cells in the periphery connect with many rods, and in turn information from a number of bipolar cells is carried to each ganglionic cell. Assuming that this simple model of structure also applies to function, a given stimulus "edge" of adjacent light and dark extents would be less likely to be processed in peripheral regions where both components strike cells that stimulate the same ganglion. In this framework one would expect size to be an increasingly important variable as angular

deviation increases. That did not occur in this experiment, probably because of floor effects upon performance at 25° and 58°.

Brightness had a significant effect on correct response performance in all regions except the extreme periphery. It was not a strong effect, however. That supports Mandelbaum and Sloan's (1947) conclusion that "light intensity is not as critical a factor in peripheral acuity as it is in central acuity [p. 587]." That was especially true with the response time measure where brightness did not have an effect after 10°.

Exposure duration increased the accuracy of performance only in the near periphery, but it did have a significant effect upon response time. As is commonly found, the longer that the *S*s had to view the stimulus, the longer they took to respond. Aside from the near periphery, however, exposure duration did not affect the level of performance.

One of the most valuable contributions of this study is its demonstration that parts of the periphery, specifically the near and middle periphery, are capable of performing more complex types of processing than the visual information theories of Gibson (1966), Mackworth (Mackworth & Morandi, 1967), and others usually ascribe to them. Perhaps instead of talking of the visual system as dichotomous in nature, our notion should be extended to 3 or more areas based upon processing capacities with no sharp dividing line between them. The first area would be the fovea, extending to perhaps 2½°. The second area would extend from just outside the fovea up to and perhaps through the near periphery at 15°. With practice in attending in the periphery, an *S* could here handle information leading to judgments of identification or classification. The third area would include the middle periphery from about 15° to about 25°, and this region appears to have a limited ability at processing complex information. The fourther is the remaining peripheral region, which may have only the crude and primitive processing capacity of which Gibson and Mackworth write.

REFERENCES

Engel, F. L. Visual conspicuity, directed attention and retinal locus. *Vision Research*, 1971, **11**, 563-576.

Gibson, J. *The senses considered as perceptual systems.* Boston: Houghton Mifflin, 1966.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Hubel, D., & Wiesel, T. Receptive fields and functional architecture in two non-striate areas of the cat. *Journal of Neurophysiology*, 1965, **29**, 229-289.

Kerr, J. L. Visual resolution in the periphery. *Perception & Psychophysics*, 1971, **9**, 375-378.

Leibowitz, H. W., Johnson, C. A., & Isabelle, E. Peripheral motion detection and refractive error. *Science*, 1972, **177**, 1207-1208.

Low, F. The peripheral visual acuity of 100 subjects. *American Journal of Physiology*, 1943, **150**, 83-88.

Low, F. The peripheral visual acuity of 55 subjects under conditions of flash presentation. *American Journal of Physiology*, 1947, **151**, 319-324.

Mackworth, N. H. Visual noise causes tunnel vision. *Psychonomic Science*, 1965, **3**, 67-68.

Mackworth, N. H., & Morandi, A. J. The gaze selects informative details with pictures. *Perception & Psychophysics*, 1967, **2**, 547-552.

Mandelbaum, J., & Sloan, L. Peripheral visual acuity. *American Journal of Ophthalmology*, 1947, **30**, 581-588.

Menzer, G. W., & Thurmond, J. B. Form identification in peripheral vision. *Perception & Psychophysics*, 1970, **8**, 205-209.

Newsome, L. R. Visual angles and apparent size of objects in peripheral vision. *Perception & Psychophysics*, 1972, **12**, 300-304.

Payne, W. H. Visual reaction times on a circle about the fovea. *Science*, 1967, **155**, 481-482.

Polyak, A. *The retina.* Chicago: University of Chicago Press, 1941.

Sanders, A. F. *The selective process in the functional visual field.* Soesterberg, The Netherlands: Institute for Perception RVO-TNO, 1963.

Sanders, A. F. Some aspects of the selective process in the functional visual field. *Ergonomics*, 1970, **13**, 101-117.

Vaughan, G., & Corballis, M. Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. *Psychological Bulletin*, 1969, **72**, 204-213.

Wertheim, T. Über die indirekte Sehschärfe. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 1894, **7**, 172-187.

Williams, L. G. The effect of target specifications on objects fixated during visual search. *Perception & Psychophysics*, 1966, **1**, 315-318.

Yarbus, A. *Eye movements and vision.* New York: Plenum Press, 1967.

(Received April 19, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 250–255

REACTION TIME TO PHONEME TARGETS AS A FUNCTION OF RHYTHMIC CUES IN CONTINUOUS SPEECH [1]

JOYCE L. SHIELDS [2]

U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, Virginia

ASTRID McHUGH AND JAMES G. MARTIN

University of Maryland

Two experiments using the same *Ss* in one sitting tested the hypothesis that the temporal redundancy inherent in rhythmically structured sound sequences is used during listening by the perceptual mechanism. Reaction times (RTs) were recorded for *Ss* who responded to the target phoneme /b/ in running speech. In the main experiment sentences containing the target phoneme as the word initial phoneme in either accented or unaccented syllables were presented to listeners. In the control experiment the words containing the target phoneme were spliced out of the sentence context and embedded in a string of nonsense words and presented to the same *Ss*. The RTs to the target phoneme were faster to accented than to unaccented targets in the sentence context, whereas RTs did not differ significantly in the nonsense sequence context. These results indicate that RT cannot be explained by simple differences in articulatory and other characteristics between accented and unaccented targets, and that temporal redundancy is a determining factor in RT to targets in connected speech. Accented targets are temporally predictable whereas unaccented targets are not.

Consideration of the rhythmic movement involved in serial behaviors led Lashley (1951) to propose a general timing mechanism underlying all natural movement sequences. Rhythmic action and hierarchical motor organization were suggested to be related concepts which might provide the natural link between the perception and production of all integrated movements, including speech. The details of such a relationship were not made clear at that time, although the need for methods to analyze such systems was emphasized. Recently, Martin (1972) has offered 2 rules for the formal description of simple natural rhythmic patterns, and has shown how rhythmic action has consequences for theories of motor and perceptual behavior in real time. The general thesis developed by Martin suggests that the organization of speech sequences as well as all natural movement sequences are constrained by relative timing (i.e., rhythm).

Martin (1972) hypothesizes that speech elements are hierarchically organized in a coherent, internal temporal structure at the sound level. That is, the locus of each sound element along the time dimension is determined relative to the locus of all other elements in the sequence, adjacent and nonadjacent. The sequence is naturally constrained by the underlying temporal organization. An alternative to this viewpoint suggests that speech elements are simply ordered in a successive (concatenated) sequence. According to the concatenation view the relationship between elements at the sound level is that of

[1] This research was supported in part by Grant MH-16726 from the National Institute of Mental Health, U.S. Public Health Service, to J. G. Martin; a grant from the Bio-Medical Sciences Support Committee, University of Maryland, to the Center for Language and Cognition; a National Science Foundation graduate traineeship to A. McHugh; and a U.S. Army long-term training fellowship to J. L. Shields, who was on sabbatical leave to the University of Maryland for the year. The major contributions to all aspects of this work were shared equally by J. L. Shields and A. McHugh. Thanks are due to John McHugh for the design and construction of the apparatus. Computer time for this project was supported in full through the facilities of the University of Maryland. The opinions expressed in this paper are not necessarily those of the Department of the Army.

[2] Requests for reprints should be sent to Joyce L. Shields, ARI, 1300 Wilson Boulevard, Arlington, Virginia 22209.

ordering, without regard to their structural relationships or organization at some other level. These alternatives, rhythmic patterning and concatenation, are related to distinct theoretical points of view concerning real-time behavior, and they have distinct implications for the perception and production of speech. The production of rhythmic speech patterns requires temporal patterning between adjacent and nonadjacent speech sounds, whereas the concatenative alternative requires only correct temporal sequence. These alternative viewpoints make distinctly different predictions about the degree of organization in speech sounds and in their perception. The rhythmic viewpoint suggests that certain elements in a speech sequence should be more predictable in time than others, whereas the concatenative viewpoint would appear to suggest that each speech element is equally predictable (or better, equally unpredictable) in time. The hypothesis generated by the rhythmic theory will be developed in the following.

Rhythmic patterns are element sequences in which some elements are accented, standing out to some degree, whereas other elements are not. The main implications of the rhythm rules stated by Martin (1972) are (*a*) accent level covaries with timing, and (*b*) main accents remain equidistant. If speech sequences are rhythmically patterned, then some of the speech elements will be accented relative to the other elements. The hierarchical organization of these temporal sequences would suggest that main accents would fall on the potentially most important elements of the speech sequence. Therefore, a logically consistent hypothesis would suggest that an efficient perceptual mechanism would make use of the information furnished by the prosodic (rhythmic) contours of the speech pattern. That is, a rhythmic pattern is temporally redundant so that once early accents are heard the location of later accents (and perhaps the end of the pattern as well) can be predicted in time. Such a perceptual mechanism would allow for the focusing of attention on the elements in the speech sequence receiving accent. The hypothesis tested in this experiment is that the temporal redundancy in the speech sequence is used by the perceptual mechanism.

Martin's (1972) theory predicts that the listener is capable of anticipating the occurrence of accented speech elements on the basis of the prosodic contours produced by the speaker. This implies that the listener's attention will be focused anticipatorily on the accented syllables (i.e., accented syllables are targets during perception). The relationship between accented syllables and relative timing is such that certain aspects of the spoken signal become temporally predictable to the listener once he has "locked in" to the pattern. Presumably the hearer's targets in listening will be the same as the targets of the speaker—namely, the accented syllables—as these are likely to be the most informative segments of the speech sequence. That is, the listener's perceptual strategy is such that during ongoing speech, it is possible for him to anticipate in real time the onset of accented targets and also the end of the speech sequence. Therefore, if he is asked to respond to a given aspect of the speech signal, for example, the initial phoneme in a syllable, his reaction time (RT) should be faster to phoneme targets in accented syllables than to phoneme targets in unaccented syllables. Since unaccented syllables are not temporally redundant with respect to the remainder of the pattern, they are not expected at any particular time.

Of course, faster RTs might also be predicted on the basis that a target phoneme in an accented syllable is not only louder, etc., but is also more clearly articulated than one in an unaccented syllable and hence should be perceived more easily. Experimental support of this point of view is provided by a study by Lieberman (1963). In that study experimental words were gated out of running speech. Lieberman found that words spoken in the context of sentences where they were informative were more easily recognized than the same words spoken in a context where they were redundant. If quality of articulation and

other characteristics of accented syllables themselves are the determining factor in RT, faster RTs would be expected to the accented target. Therefore, an experimental test of Martin's (1972) theory must demonstrate that the results cannot be explained by simple differences in articulatory and other characteristics between accented and unaccented targets.

Two separate experiments were conducted using the same set of *S*s. That is, each *S* participated first in the main experiment and then in the control experiment in one sitting. In the main experiment sentences containing the target phoneme /b/ as the word initial phoneme in accented and unaccented syllables were presented to listeners. The *S* was asked to listen for the phoneme and respond by pressing a button as rapidly as possible. In the control experiment the words containing the target phonemes had been spliced out of the running speech and embedded in a string of nonsense words, and *S* was asked to perform the same task. If temporal redundancy is a determining factor in RT, RTs to the accented targets should be faster than RTs to the unaccented targets in the sentence context, and to a greater extent than in the control experiment.

Method

Design

Each *S* served both in the main and control experiments. First he listened to a tape, which consisted of 30 sentences. His task was to press a button as quickly as possible whenever he heard the phoneme /b/ and to write down the /b/ word after the sentence ended. For the main (sentence context) experiment the design was a 2 × 3 repeated measurements design (accented or unaccented target and 3 phrase locations). For the main vs. control experiment comparison, *S* listened to 30 sequences, each consisting of 8 2-syllable nonsense words, one of which was the /b/ word, and performed the same task as for the sentences. The design was a 2 × 2 repeated measurements design (accented or unaccented target and nonsense or sentence context). There were 6 different tapes, each using the same sentence frames and nonsense sequences. All conditions were represented on each tape, but the particular target conditions and words occurred in different sentences and nonsense contexts on the different tapes.

Materials

Main experiment (sentence context). The experimental list consisted of 30 sentence frames and 30 2-syllable nonsense words which were used in the sentences in a context where a noun would normally occur. The nonsense word occurred at the beginning, middle, or end of the last phrase of each sentence. These phrases followed a natural clause break in the sentence and were approximately 7–9 syllables long. Across the 6 tapes the nonsense word occurred equally often at the beginning, middle, or end of the phrase. An example of each location follows.

Beginning. They had long decided that they were totally lost in the maze of side streets, for BENkik was a city of alleys.

Middle. You will hav[illegible]il any morning sightseeing plans, as the plane to benKIK leaves at noon.

End. First the group hiked along the side of the stream, then they forded the BENkik.

Twenty-four of the sentences had nonsense words that began with the phoneme /b/ (/b/ words). The nonsense words in the remaining 6 were distractors. The sentences were designed so that the phoneme /b/ would occur only in the target words. In the course of the experiment, it was discovered that 2 of the sentences each had another word with /b/ in it, e.g., "by," in the clause prior to the experimental phrase. Only a few *S*s responded to the accidental /b/s. These responses could be easily identified and were eliminated from the data. Because of the manner in which nonsense words and syllable accents were balanced across tapes, the accidental /b/s were balanced across conditions.

The 6 tapes were constructed using the 30 sentence frames and target words once each on any given tape. Balanced across tapes, each nonsense word occurred in all 6 conditions. Across the 6 experimental tapes each word containing a target was placed in 3 different sentence frames so that each word could occur in each phrase location—beginning, middle, and end. In each phrase location and sentence the target syllable was either accented or unaccented. For the example given above, on one tape the word Benkik would be accented on the target syllable (i.e., BENkik) and on another tape it would occur in the same context but with the target syllable unaccented (i.e., benKIK). Within each tape each of the 6 possible conditions (3 phrase locations and accented vs. unaccented) occurred 4 times with /b/ words and once each with distractors.

An experienced female reader recorded each of the 6 experimental sentence lists on 1 channel of a 2-channel tape. The sentences were read with natural emphasis and intonation. The appropriate accent for the nonsense word was marked on the typed material from which she read.[2] The record-

[2] This procedure, by drawing attention to the syllable to be accented, led the speaker to give the syllable primary accent. Thus it enhances the differences between accented and unaccented

ings were checked for pronunciation and accent by *E*s and those found inadequate were rerecorded. The reader was naive as to the theoretical basis for the experiment.

A tone was recorded on the second channel of the tapes, starting at the beginning of each phoneme target (the tone could not be heard by *S*). The tone was used to close a relay which was in series with 4 electronic timers and 4 buttons. The timers started when the tone operated the relay. Each timer was separately connected to a button and when *S* pressed the button the timer was stopped. For the purpose of recording the tone on the second channel, the beginning of the phoneme target was located by moving the tape manually back and forth across the playback head. There were probably slight variations in the alignment of tone onset and the beginning of the nonsense word, but such variations were probably randomly distributed. Two copies were made of each of the original sentence tapes (with tones). One of these copies served as the experimental tape in the main experiment.

Control experiment (nonsense word sequence). The nonsense words used as phoneme targets and the distractor nonsense words were cut out of the second copy of the sentence tapes. Thirty sequences of 7 2-syllable nonsense words were recorded by the same female reader. These words were then cut out, scrambled, and respliced so that there was no natural pattern to the sequences. Using a predetermined nonsystematic order, the 24 /b/ words and 6 distractors from the sentences were spliced into the nonsense sequences, occurring equally often as the third, fifth, or seventh item in the sequence, which now had 8 items. The tone on the second channel was included, so for a given word the timer would be started at exactly the same time in both the sentence and the nonsense context. These nonsense word sequences were used for the control experiment.

Subjects

The *S*s were 48 students from introductory psychology classes at the University of Maryland who volunteered to participate for extra credit. Eight *S*s were assigned to each of the 6 experimental tapes.

Procedure

The *S*s were tested in groups of 1–4, with *E* and *S* occupying adjoining rooms. Each *S* was seated out of direct view of the others and so that his arm and hand could rest comfortably with the preferred hand resting 3 in. in front of the response button.

General instructions outlining *S*'s task were recorded on the experimental tape and played at the beginning of the experimental session. The *S*s were told to listen for the sound "b as in boy" and to press the button as quickly as possible when they heard it. In addition, they were told to write down the word after the sentence ended. This instruction insured that *S* was attending to the sentences, and at the same time it provided a check on what he had responded to.

At the end of the taped instructions, *E* returned to the *S* room to answer questions, check hand positions, and clarify the instructions. The *S* was familiarized with the RT task through a series of 4 practice sentences, 3 of which contained a /b/ word. Each *S* listened first to the 30 sentences and afterward to the 30 sequences of nonsense words played over earphones at a comfortable listening level. The *E* recorded individual RTs for each sentence on prepared recording sheets.

targets, since if the second syllable is strongly accented, then (everything else being equal) the first (target) syllable naturally receives minimal accent according to the rhythm rule (Martin, 1972).

Results

Main Experiment (Sentence Context)

Six median RTs were obtained for each *S*, one for each Accent × Phrase Location condition. Table 1 shows the mean scores for all *S*s for these 6 conditions. A Phrase Location × Accent analysis of variance revealed that *S*s responded faster to accented targets than to unaccented targets at all phrase locations, $F(1, 47) = 49.60$, $p \leq .001$. At the 3 locations, *S*s responded more quickly the later the target appeared in the phrase, $F(2, 94) = 19.92$, $p \leq .001$. The significant main effects must be interpreted with reference to the significant interaction of phrase location and accent, $F(2, 94) = 8.00$, $p \leq .001$. Differences between pairs of means were tested by the Tukey procedure (Winer, 1971). The RTs to accented targets were significantly faster than RTs to unaccented targets in the early and middle phrase locations. For accented targets, RTs were significantly faster to targets in the middle and late

TABLE 1

Mean Reaction Time Scores (in msec.) for Accent × Phrase Location Conditions

Accent	Phrase location			
	Beginning	Middle	End	$\bar{X}$
Accented	594	537	547	559
Unaccented	657	640	562	619
$\bar{X}$	625	589	554	589

TABLE 2

MEAN REACTION TIME SCORES (IN MSEC.) FOR ACCENT × CONTEXT CONDITIONS

	Context		
	Sentence	Word sequence	
Accented	549	671	610
Unaccented	612	687	650
$\bar{X}$	581	679	630

phrase locations than to targets in early phrase locations. For unaccented targets, RTs were significantly faster to targets in the late phrase location than to the early or middle phrase locations.

Control Experiment (Nonsense Sequence)

Four separate median RTs were calculated for each *S* for the accented targets and unaccented targets in the sentence context and the nonsense sequence context. The mean scores for all *Ss* for these 4 conditions are shown in Table 2. An analysis of variance showed *Ss* responded faster to accented targets than to unaccented targets, F (1, 47) = 21.28, $p \leq .001$, and faster to targets in the sentence context, F (1, 47) = 34.10, $p \leq .001$, than in the nonsense sequence context. In addition, there was a significant Context and Accent Conditions interaction, F (1, 47) = 5.80, $p \leq .025$. Tests of the individual mean comparisons by the Tukey procedure (Winer, 1971) indicated that the sentence context RTs were shorter when the targets were accented than when the targets were unaccented. But when these same targets were spliced out of the sentence context and spliced into the nonsense word sequences, the differences in RTs to accented and unaccented target phonemes were not significant.

DISCUSSION

The results obtained in the present study suggest that the listener uses the rhythmic ...ttern of speech to anticipate accented ... in the speech sequence. The most ...nt comparison in support of such an interpretation concerns the target locations in the sentence context. In the sentence context the sharp drop in RT to accented targets in the middle location as shown in Table 1 is to be expected on the theory that some minimal temporal context (e.g., 2 "beats") must occur to establish the rhythm of a sequence. Only then do the anticipatory consequences of temporal redundancy become useful to the listener. That is, the listener's attention is focused on the primary accents in the sequence and therefore RTs to targets occurring coincident with the primary accent would be expected to be faster than to those targets which do not receive primary accent. For the unaccented syllables, the listener's attention will focus more strongly on the accented syllable which follows and it should take him longer to recognize the target in the unaccented syllable. Note that this does not imply that the listener is aware of the temporal redundancy or is capable of making deliberate use of it, as the reader can check by listening momentarily to someone speaking in the hallway. Rather, the argument here is that since syntactic and semantic context (redundancy) is ruled out, the results of the experiment can be taken as evidence that the postulated temporal redundancy exists in natural running speech.

However, in the light of the above argument, it would be predicted that RTs to accented targets should improve toward the end of the phrase, but why should RTs to unaccented targets also improve? On the present th... rhythmic patterns, here phrases, are ho... internally organized units so that the listener can anticipate the end of the unit without a terminal cue. Hence, the faster RTs at the terminal phrase locations may be due, in part, to anticipatory guessing based on rhythmic cues. Redundancy due to temporal as well as other factors makes the targets more predictable toward the end of the sentence. Support for this interpretation comes from other published work. Several phoneme-monitoring experiments (Cairns & Foss, 1971; Foss, 1969; Foss & Lynch, 1969; Hakes, 1972; Hakes & Cairns, 1970; Hakes & Foss, 1970) were run to show that latency differences depended upon differential difficulty levels in the processing of linguistic inputs. However, with the exception of one early experiment (Foss, 1969), in all studies where early and late phoneme positions were contrasted, RTs to the phoneme in the late position were found to be faster than to the phoneme in the earlier

position, which is in agreement with the results obtained in the present study, and at variance with the hypothesis that the burden on short-term memory increases as processing proceeds [illegible]

In addition, the differences observed in the present study cannot be explained by theories of linguistic complexity, since the sentence contexts were the same for accented and unaccented syllables. Therefore, differences in monitoring must be due to differences at the [illegible]tic level. These differences must be more than differences in articulation of the target phoneme since the differences between accented and unaccented syllables were nonsignificant in the nonsense sequence context. It is consistent with the present results to hypothesize that much that has been left unexplained by theories of linguistic complexity may be clarified on the grounds of temporal patterning.

Concerning the comparison between sentences and nonsense sequence contexts, when a word beginning with an accented target phoneme was spoken in the rhythmical context of the sentence the listener's RTs were significantly faster than when the same word occurred in a nonsense word sequence. In the rhythmic context of the sentence, Ss were able to respond more quickly to the accented targets than to the unaccented targets. Yet when exactly the same acoustic signal was spliced out of the sentence context and placed among other isolated nonsense words, accented and unaccented target phonemes did not produce significant differences in RT. It would appear that the rhythmic context of the sentence increased the predictability of the accented target phonemes. The differences in RTs when the targets were in the sentence context vs. the list of nonsense words cannot [illegible] word itself, since the words were spliced out of the sentence and embedded in the nonsense word context. While quality of articulation may be a factor in RT, the acoustic signal was [illegible]

In the nonsense sequence there is no rhythm[illegible] context the listener can use to anticipa[illegible] accented targets, so attention cannot be focused in advance on the accented target. Hence, the advantage of the accented over unaccented targets diminishes and the u[illegible] predictability results in generally longer R[illegible]

REFERENCES

[illegible]

[illegible], J. [illegible] comprehension: Effects of lexical item difficulty and position upon decision times. *Journal [illegible] Verbal Learning and Verbal Behavior*, 1969, [illegible] 457-462.

Foss, D. J., & Lynch, R. H., Jr. Decision processes during sentence comprehension: Effects of surface structure on decision times. *Perception & Psychophysics*, 1969, 5, 145-148.

Hakes, D. T. Effects of reducing complement constructions on sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 278-286.

Hakes, D. T., & Cairns, H. S. Sentence comprehension and relative pronouns. *Perception & Psychophysics*, 1970, 8, 5-8.

Hakes, D. T., & Foss, D. J. Decision processes [illegible] comprehension: Effects of surface [illegible]

[illegible]

Lieberman, P. Some effects of semantic [illegible] grammatical context on the production [illegible] perception of speech. *Language and Speech*, [illegible] 6, 172-179.

Martin, J. G. [illegible] serial structure in speech [illegible] *Psychological Review*, 1972, 79, 487-509.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

(Received April 24, 1973.)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 256–259

EFFECTS OF BASELINE SELF-REINFORCEMENT BEHAVIOR AND TRAINING LEVEL ON POSTTRAINING SELF-REINFORCEMENT BEHAVIOR

ALBERT KOZMA[1]

Memorial University of Newfoundland, St. John's, Newfoundland, Canada

PAMELA EASTERBROOK

University of New Brunswick, Fredericton, New Brunswick, Canada

The present investigation evaluated the effects of baseline self-reinforcement scores and level of training on posttraining self-reinforcement behavior. The *S*s were classified as high, medium, or low self-reinforcers on a pretraining task. An equal number of *S*s from each baseline grouping were assigned to one of 3 training levels and trained to 40%, 60%, or 80% criteria on a 4-choice discrimination learning task. The training phase was followed by a posttraining phase during which rates of self-reinforcement behavior were assessed. Results showed that training raised baseline self-reinforcement behavior only if level of training exceeded baseline self-reinforcement rates. Moreover, the 3 baseline groupings differed in the number of correct and incorrect application of self-reinforcement scores.

In recent years, considerable effort has been exerted in an attempt to isolate the variables affecting self-reinforcement (SR) behavior (Bandura & Whalen, 1966; Haynes & Kanfer, 1971; Skinner, 1953). Investigators using the directed learning paradigm to assess SR behavior have frequently concluded that *S*s tend to match SR rates at which external feedback had been provided during training Kanfer & Duerfeldt, 1967; Kanfer & Marston, 1963). A closer inspection of the literature strongly indicates that such a conclusion is an oversimplification. In studies in which the level of training was *low* (where the criterion level was less than 5 correct responses in 10 trials on the last block of training trials, or where noncontingent reinforcement was administered on fewer than 50% of the training trials), SR scores during the test phase *exceeded* training levels (Bartol & Duerfeldt, 1970; Kanfer & Marston, 1963; Marston, 1969).

These data suggest that increases in SR behavior will occur only if training rates exceed baseline SR scores. One of the aims of the current investigation was to test the preceding conclusion.

A second aim of the current investigation was to determine whether a sufficiently low training level would significantly lower the number of SR responses of high baseline (HB) *S*s. In the one relevant study reported in the literature (Bartol & Duerfeldt, 1970), a low training level had no effect on the somewhat higher SR baseline rate.

A third concern of the present investigation was to determine whether an SR change following training is indiscriminate, or whether it involves an increase in the accuracy of SR/CSR (self-reinforcement on correct choices) matching. If SR responses become highly correlated with correct choices as a result of training, SR rates may merely reflect the degree of certainty *S* has in the accuracy of his choice, rather than reflecting some nebulous internal motivating state.

METHOD

Subjects. The *S*s were 45 male and 45 female introductory psychology students attending Memorial University of Newfoundland. Each *S* was paid $1.00 per half hour of experimental participation.

Apparatus and materials. Sixty 24 × 36 mm. slides, each showing a consonant–vowel–consonant nonsense syllable in every quadrant, were used as task stimuli. The 60 slides constituted 6 copies of 10 basic slides. The association values of the nonsense syllables were 49%–51% (Archer, 1960). The type of syllable, its position on the slide, and the order of slide presentation were randomly determined.

[1] Requests for reprints should be sent to A. Kozma, Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

Stimuli were presented with a Kodak Carousel 800 slide projector and covered an area of 17.8 × 27.9 cm. on the screen. The 43.2 × 25.4 × 7.6 cm. response panel had 4 toggle switches mounted along its base on S's side. The switches, numbered 1, 2, 3, and 4, represented the 4 different quadrants of the projected image. A diagram above each switch denoted the relevant quadrant. A red SR button was centered above the switches, and a white reinforcement light (lamp size 1892) activated by the SR button was mounted above it. The E's side of the panel was identical to S's side, except that 4 differently colored indicator lights replaced the 4 toggle switches. To control the 4-sec. exposure time, the 4-sec. response interval, and the 10-sec. interlock interval, BRS logic units were used.

Design and procedure. An equal number of HB, medium baseline (MB), and low baseline (LB) self-reinforcers of both sexes received 40%, 60%, or 80% task contingent training. This procedure resulted in 18 treatment cells of 5 Ss per cell.

The experiment involved 3 distinct phases—baseline, training, and posttraining or test phases. Immediately preceding the baseline phase, Ss were shown their side of the response panel. They were presented with the 4-choice discrimination task, and instructed to select the correct syllable by pressing the switch which corresponded to the quadrant in which their choice was located. The task was presented as a forced-choice task, and Ss were required to give their best choice on every trial, even if that choice was only a guess. In addition, Ss were asked to press the SR button whenever they thought that their selection of a nonsense syllable had been correct. All Ss received 3 demonstration trials to familiarize themselves with the equipment and the task.

The 3 demonstration trials were followed by the 20 trials of the baseline phase. Feedback was entirely absent during this phase, and the probability of making a correct choice was at chance level ($p = .25$). The Ss with SR scores in the upper third of the distribution were labeled HB, those with scores in the middle third were labeled MB, and those with scores in the lower third were labeled LB. The Ss with similar SR scores during the baseline phase were randomly assigned to the training conditions. Separate distributions of baseline SR scores were calculated for males and females.

Immediately following the baseline phase, Ss were instructed not to push the SR button, since E would turn on the reinforcer light every time they made a correct choice. The training phase, with feedback for correct choices, was continued until Ss reached their criterion level of 4, 6, or 8 correct choices on a block of 10 trials.

The training phase was followed by a 20-trial test phase. Before this final phase, Ss were once more asked to take over the reinforcer role and to provide their own reinforcement for correct choices by pushing the SR button.

The Ss' task responses were recorded for each

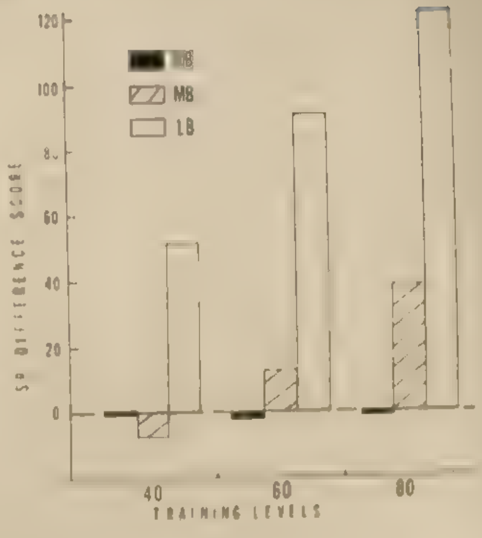


FIGURE 1. Representation of the Baseline Grouping × Training Level interaction. (SR = self-reinforcement; HB = high baseline; MB = medium baseline; LB = low baseline.)

phase of the study while, in addition, SR responses were taken during baseline and test phases.

RESULTS

The mean number of blocks of trials to criterion during training was 3.9, 6.4, and 8.9, respectively, for 40%, 60%, and 80% criterion groups. Neither baseline grouping nor sex had a differential effect on training rate.

A preliminary analysis of baseline SR scores indicated that base rates for males in the LB condition were significantly higher than for females ($p < .01$). Within each sex group, baseline SR scores did not, however, differ for 40%, 60%, and 80% training groups.

SR change. To assess the effectiveness of the different training levels in modifying SR behavior, SR change from pre- to posttraining was evaluated. The SR difference scores were analyzed by a 3 × 3 × 2 (Baseline Grouping × Training Level × Sex) analysis of variance design. Significant effects of baseline, $F(2, 72) = 78.14$, $p < .01$, training, $F(2, 72) = 13.03$, $p < .01$, and the Baseline × Training interaction, $F(4, 72) = 3.98$, $p < .01$, were obtained.

TABLE 1

MEAN BASELINE RATIO OF SELF-REINFORCEMENT (SR) RESPONSES FOLLOWING CORRECT CHOICES (CSR) TO CORRECT RESPONSES (CR) AND OF SR RESPONSES FOLLOWING INCORRECT CHOICES (ISR) TO INCORRECT CHOICES (IR), AND EXPECTED AND OBTAINED POSTTRAINING SR SCORES

Baseline		Training level											
		80%				60%				40%			
Grouping	SR score	CSR/CR	ISR/IR	SR predicted	SR obtained	CSR/CR	ISR/IR	SR predicted	SR obtained	CSR/CR	ISR/IR	SR predicted	SR obtained
High	98	.99	.90	98	96	97	.91	94	95	.97	.95	96	98
Medium	63	.93	.32	80	80*	.90	.25	64	66	.67	49	56	58
Low	13	.81	21	69	75*	.83	.40	65	61*	.51	36	[illegible]	42*

* $p < .05$ between SR baseline and posttraining scores.

An inspection of the treatment means contributing to the significant effects (Figure 1) suggests that there was little change in SR behavior of HB *Ss*, while a substantial increase in SR behavior of LB *Ss* appeared to have occurred at all training levels. Other relationships, especially the effects of training on MB *Ss*, were less clear. Accordingly, a multiple comparison of the relevant treatment groups was undertaken by means of the Newman-Keuls procedure (Winer, 1962).

Results of the comparison showed that LB *Ss* increased their SR behavior significantly more than MB and HB *Ss* after 40% and 60% training ($p < .05$). After 80% training both LB and MB *Ss* showed a significant increase over HB *Ss* ($p < .05$), with significantly greater changes occurring in the former than in the latter group ($p < .05$). Since SR difference scores for HB *Ss* do not reflect a significant pretraining–posttraining change ($p > .10$), it may be concluded that an increase in SR behavior occurred for LB *Ss* at all levels of training and for MB *Ss* only after 80% training. Training did not appear to modify SR behavior of HB *Ss* at any level. An inspection of "SR score" and "SR obtained" columns of Table 1 shows that training led to an SR change only when it exceeded SR baseline performance.

SR accuracy. A second concern of the current investigation was to determine whether increases in SR behavior were indiscriminate, or whether they reflected an increased accuracy in SR – correct response (CR) matching. To resolve the issue, postraining SR responses were divided into those which followed correct choices (CSR) and those which followed incorrect choices (ISR). The CSR/CR ratio is a good indicator of omission errors (i.e., the failure to provide self-reinforcement after making a correct response). A second type of error, the tendency to provide SR responses after incorrect choices (IR) may be represented by the ISR/IR ratio. For close SR/CR matching, CSR/CR approaches 1, and ISR/IR approaches 0.

Since HB *Ss* failed to show a significant SR change and provided self-reinforcement after virtually every response, there appeared to be little value in including them in the analysis. Both CSR/CR and ISR/IR would approach 1 (actual values: CSR/CR = .98, and ISR/IR = .92.) Training did not seem to modify SR behavior of these *Ss*.

Posttraining CSR/CR values for baseline groupings that did show an SR change with training were analyzed by a 2 × 3 × 2 (Baseline Grouping × Training Level × Sex) analysis of variance design. The only significant effects were baseline, F (1, 48) = 5.63, $p < .01$, and training, F (2, 48) = 15.53, $p < .01$. The baseline effect indicates that MB *Ss* had significantly fewer omission errors than LB *Ss* (CSR/CR = .83 and .72, respectively).

A comparison of training means showed

that the 80% and 60% groups did not differ from each other, while the 40% group differed significantly from both, $p < .05$ (CSR/CR = .87, .86, and .69, respectively). There were, thus, significantly more omission errors after 40% training than after 60% and 80% training.

Posttraining ISR/IR values were also analyzed by a $2 \times 3 \times 2$ (Baseline Grouping $\times$ Training Level $\times$ Sex) analysis of variance design. None of the effects of this analysis were significant ($p > .05$). Although commission errors did occur for MB and LB *S*s, they occurred considerably less frequently than for HB *S*s, $p < .001$ (ISR/IR = .32, .34, and .92, respectively).

The CSR/CR and ISR/IR values for baseline groupings and training levels, together with the number of SR responses predicted by these values, and the number of SR responses actually obtained, are all recorded in Table 1. The table shows that CSR/CR and ISR/IR ratios together are good predictors of SR behavior following training.

Discussion

The data on SR changes were consistent with the prediction that training modifies SR behavior whenever training level exceeds baseline SR performance. They clearly illustrate that posttraining SR rate is not solely a function of training, but is affected by *S*'s initial operant SR rate. Results strongly suggest that posttraining SR responses reflect the amount of confidence *S*s have in the accuracy of their response. The HB *S*s appeared to be extremely confident, and they did not modify their SR behavior as information about accuracy was increased by training. The approach lead to very few omissions but a large number of commission errors. On the other hand, LB *S*s, who appeared to be less confident, significantly increased their SR responses as training provided information about the accuracy of their responses. The commission errors were low during baseline performance and did not increase with training. The MB group appears to fall somewhere between the 2 extremes. Like LB *S*s, they increased the accuracy of their SR responses with training, but were not quite as cautious and had fewer omission errors. The MB *S*s were, therefore, the most accurate of the 3 baseline groupings.

The interpretation of SR responses as an expression of confidence in the accuracy of one's choice would also appear to account for sex differences in baseline SR scores. In a male-dominated society such as ours, it may be more difficult for males than for females to appear indecisive.

A comparison of current findings with those reported by Dorsey, Kanfer, and Duerfeldt (1971) points to a major difference in SR behavior if a contingent as opposed to a noncontingent feedback procedure is employed during training. While both procedures led to increases in CSRs with increases in external feedback, increases in ISRs with increases in external feedback occurred only when the noncontingent feedback technique was used. It would, therefore, appear that *S*s will provide self-reinforcement even if uncertain about the accuracy of a specific response, provided that there is a reasonable chance of being correct by guessing (under high levels of noncontingent feedback).

References

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74**(10, Whole No. 497).

Bandura, A., & Whalen, C. K. The influence of antecedent reinforcement and divergent modeling cues on patterns of self-reward. *Journal of Personality and Social Psychology*, 1966, **3**, 373–382.

Bartol, G., & Duerfeldt, P. Self-reinforcing behavior: The effects of base rate and dependency. *Journal of General Psychology*, 1970, **83**, 151–161.

Dorsey, T. E., Kanfer, F. H., & Duerfeldt, P. H. Task difficulty and non-contingent reinforcement schedules as factors in self-reinforcement. *Journal of General Psychology*, 1971, **84**, 323–334.

Haynes, L. E., & Kanfer, F. H. Academic rank, task feedback and self-reinforcement in children. *Psychological Reports*, 1971, **28**, 967–974.

Kanfer, F. H., & Duerfeldt, P. H. Motivational properties of self-reinforcement. *Perceptual and Motor Skills*, 1967, **25**, 237–246.

Kanfer, F., & Marston, A. Determinants of self-reinforcement in human learning. *Journal of Experimental Psychology*, 1963, **66**, 245–254.

Marston, A. Effect of external feedback on the rate of positive self-reinforcement. *Journal of Experimental Psychology*, 1969, **80**, 175–179.

Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received April 28, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 260–265

INDUCED CHUNKING:

TEMPORAL ASPECTS OF STORAGE AND RETRIEVAL[1]

CHARLEY McCAULEY AND GEORGE KELLAS

University of Kansas

Instructions to 3 treatment groups were to form groupings of "[illegible]ed" words via imagery, sentence formation, or repetition of unit members. A control group was given only typical free-recall instructions. [illegible] measures were employed that allowed the temporal flow of information to be monitored during both the storage and retrieval of list items. The results indicated that functional higher order units were formed in all instructed conditions and that the use of specific relational organizers differentially [illegible] the amount of encoding time required for chunk processing. A temporal analysis of retrieval revealed the presence of higher and lower order units in the recall sequences. Response times between induced chunk boundaries increased linearly across output positions in all conditions and were strongly influenced by the nature of the encoding strategy. The results are discussed in terms of several retrieval models.

An established tenet of modern theories of memory organization is that *S*s form higher order units to overcome limitations in the information-processing system. The formation of these units can be based on perceptual factors (Bower & Winzenz, 1969), semantic and/or syntactic relationships (Mandler, 1967), or other more elaborate relational organizers, such as interactive imagery (Paivio, 1971). However, there is little if any evidence concerning the differential effectiveness of these relational devices.

One question of primary interest concerns the ease with which units may be encoded and their subsequent stability during output. By employing measures that monitor the temporal flow of information through the processing system, it should be possible to examine the processes involved in unit encoding and subsequent utilization during recall. Recent evidence by Laughery and Schmidt (1970) and Kellas and Butterfield (1971) has indicated that the profiles of study times across input positions manifest distinct patterns that are related to rehearsal activity. These patternings are sensitive to manipulations involving retrieval plans and the formation of higher order memory units (Kellas, Ashcraft, Johnson, & Needham, in press). It was hypothesized in the present study that if *S*s could be induced to use particular relational organizers for the construction of higher order units, the differential ease of unit encoding should be reflected in input exposure times. Using similar logic, the patterning of interresponse times (IRTs) during output should allow an assessment of the extent to which the higher order units served as functional memory units. Patterson, Meltzer, and Mandler (1971), for example, have shown that longer pauses occur at unit boundaries than between items within a unit during recall. Between-group IRTs have been interpreted as representative of a search strategy employed by *S*s to retrieve higher order units. Patterson et al. (1971) postulated that the between-group times are a function of 3 components: (*a*) group exit time, (*b*) group access time, and (*c*) word access time. They suggest that group exit time and word access time remain constant

[1] This investigation was partially supported by U.S. Public Health Service Grants HD-00870 and 00183 from the National Institute of Child Health and Human Development and it formed the basis of a thesis submitted by the first author to the University of Kansas in partial fulfillment of the requirements for the master's degree. Special thanks are due to James F. Juola for helpful comments on earlier drafts of this manuscript.

[2] Requests for reprints should be sent to George A. Kellas, Department of Psychology, University of Kansas, Lawrence, Kansas 66044.

over successive units recalled, and that increases in between-group times across output positions are due only to changes in group access time. The group access times are in turn assumed to reflect an underlying search process for previously unretrieved units.

In particular, the exponential growth of between-group IRTs across successive output positions has been related to a probabilistic search model (Patterson et al., 1971; Pollio, Richards, & Lucas, 1969) in which *S* is assumed to hold representations of all higher order units in memory. During recall, higher order units are randomly sampled with replacement and subjected to a recognition test to determine if the components of a group have previously been recalled. The exponential growth function actually obtained agrees well with expectations derived from the probabilistic search model. An alternative *fixed*-search model has also been proposed by Pollio et al. Here, it is assumed that *S* stores a fixed sequence of higher order units which are systematically retrieved in serial order during recall. This model also assumes a test phase, in which *S* checks to determine if the retrieved unit has previously been produced. The major difference in predictions between the models involves the latency function describing between-group IRTs across output positions. The probabilistic model would predict an exponential increase, whereas the fixed model is best accommodated by a linear increase in between-group IRTs resulting from the serial component of a fixed search. Existing research does not allow a clear choice of any single model, since Kellas et al. (in press) have recently reported a linear increase in between-group IRTs.

In the present study, an attempt was made to determine whether temporal patterning and search processes comparable to those obtained with categorized materials are involved in the recall of higher order units formed intraexperimentally. To the extent that induced chunking does lead to the formation of functional higher order units, one would expect an interaction of type of IRT (within vs. between group) with output positions, consistent with previous research. The interaction should result from an increase in between-group IRTs across output positions, coupled with little if any increase in within-group times. It should then be possible, by examining the within-group IRTs, to evaluate the cohesiveness of the higher order units, that is, the shorter the latency between responses within a group, the more cohesive the memory unit.

In the present experiment, *S*s were instructed to form higher order units via imagery, sentences, or rote repetition. An uninstructed control group was included to provide a baseline condition against which the effects of induced chunking could be compared. Dependent measures included input and output times as well as response accuracy in order to obtain a more comprehensive view of organizational processes.

Method

Subjects. The *S*s were 60 undergraduate students enrolled in an introductory psychology course who were given class credit for participating in the experiment.

Apparatus. A Sawyer Rotomatic slide projector was used to present the stimuli on a daylight screen. A response button was programmed to the projector so that *S* could control the duration of slide presentation. The latency between the button response and visual presentation was .4 ± .05 sec. The duration of each stimulus presentation was automatically recorded by means of a pulse-stream generator, a parallel-entry control panel, and a Sodeco printout counter. Times were recorded to the nearest .05 sec. The stimuli were typed onto transparency paper and fitted into Easymount 35-mm. frames. A Lafayette Model 810 tape recorder and a lavaliere microphone were used to record *S*s' oral recall. These taped recalls were later channeled into a Mechanics for Electronics Model AHA-20 heat-writing graphic recorder to measure interword response times (IRTs). The *S*s' recall activated a pen that indicated word onset and offset on paper moving at a rate of 25 mm/sec. All IRTs were calculated to include only time between words, and word-production time was excluded from this measure (cf. Pollio et al., 1969).

Lists. The stimulus materials were 144 nouns with high concreteness, meaningfulness, and imaginal values in the Paivio, Yuille, and Madigan (1968) word norms. In addition, all words were rated A or AA in frequency (Thorndike & Lorge, 1944). From this pool, 12 lists of 12 items each were randomly constructed with the restriction that no obvious associations existed within any list. Lists

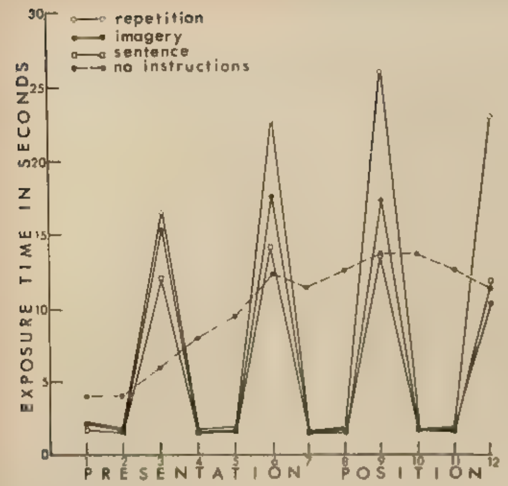


FIG. 1. Mean exposure time for items at each presentation position for all treatment groups.

were equated for concreteness (M = 6.75, SD = .099), meaningfulness (M = 6.94, SD = .146), and imagery value (M = 6.42, SD = .039).

Design and procedure. The *Ss* were randomly assigned to 1 of 4 instruction conditions, yielding a total of 15 *Ss* in each condition. The *Ss* in the no instruction (NI) condition were treated as a control group, and received only standard free-recall instructions. In addition to free-recall instructions, *Ss* in the imagery (I), sentence (S), and repetition (R) conditions were told to learn the lists by grouping items into 3s and forming an interactive image, forming a sentence, or repeating the grouped items contiguously, depending upon their treatment condition. The *Ss* in the I, S, and R conditions were told to expose the first 2 words in each successive group of 3 items only long enough to read them and to then spend as much time as they wished on the third word of each group. All *Ss* were told that item presentation would be self-paced, and that a red field would signal the end of each list and the beginning of the recall period.

The *Ss* were tested individually and received one study and one test trial on each of the 12 lists. The first 2 lists were practice lists and were not included in subsequent analyses. The recall period was unlimited, and *Ss'* oral recall was recorded on magnetic tape for further analysis. Following recall of the final experimental list, all *Ss* were given a 2-page questionnaire. On the first page, they were asked to describe in their own words the method or methods used to learn the items. Page 2 contained a list of possible encoding strategies, including imagery, sentence formation, repetition, and cumulative rehearsal, together with appropriate examples. The *Ss* were to indicate on 9-point rating scales the extent to which they employed each of the suggested strategies.

RESULTS

The major focus of the present research was on the temporal correlates of information processing during storage and retrieval. To the extent that performance approaches error-free responding, time data should provide a relatively pure index of storage and retrieval activities. Analyses pertaining to input processing, recall accuracy, and output processing are presented successively.

Input time. Figure 1 shows the exposure-time profiles across presentation positions for all treatment conditions. Exposure times for the first 2 input positions of each triad are nearly invariant for the instructed conditions, with a mean of about 2 sec. This indicates that *Ss* in these conditions were in fact following the instructions given them to spend only enough time at these positions to read the items. The processing time for each triad, as well as any time spent rehearsing previous triads, is reflected by the sharp peaks at the terminal item of each group.

The mean exposure times for each group of 3 items were computed for all conditions. Whereas these means reflect differences obtained only for the terminal items of each triad for the instructed conditions, they allow comparisons to be made between these groups and the NI condition in which *Ss* were free to spend as much time as they wished at all presentation positions. A 4 × 4 mixed analysis of variance with factors corresponding to instructions (NI, I, S, and R) and presentation position (1–4) was then performed on the mean exposure times for the triads. The main effects of instructions, F (3, 56) = 4.81, $p < .01$, and presentation position, F (3, 168) = 13.26, $p < .01$, were reliable, as was their interaction, F (9, 168) = 3.35, $p < .01$. Mean total input times for each list in the NI, R, I, and S conditions were 121, 104, 75, and 65 sec., respectively. Tukey's (b) multiple comparisons indicated that both the NI and R groups spent significantly more time during input than either I or S groups ($p < .05$). The Instructions × Presentation Position interaction resulted from the relatively large

increase in processing time across positions for the R and NI groups, as compared with relatively constant times in both the I and S conditions. The slight decrease in processing time on the last triad for all groups appeared to be related to *Ss*' tendency to output terminal items first.

Correct responding. A 4 × 10 mixed analysis of variance was conducted on the total correct responses per list in order to examine instruction effects and to ensure that no List × Instruction interaction was present.

The factors corresponded to instructions and experimental lists (1–10). The only significant finding was a lists main effect, F (9, 504) = 5.92, $p < .001$, which resulted from improved recall across lists in all conditions. Mean percent correct responses in the NI, R, S, and I conditions were 84, 80, 76, and 82, respectively.

In previously reported research under *E*-paced conditions, imagery and sentence instructions have typically led to superior performance relative to repetition and uninstructed control groups. It has been suggested (Paivio, 1971) that this recall facilitation may be related to the ease of unit encoding. The results of the present investigation confirm this hypothesis. It is clear that encoding materials in imaginal or sentential form greatly decreased the amount of processing time required by the I and S groups to reach a level of correct responding equal to that of the R and NI groups.

A further analysis of the recall data was conducted to determine the extent to which items grouped during input were recalled contiguously. Since the nominal items which comprise each higher order unit or chunk are known in the instructed conditions, the mean chunk size was computed across output positions for each successive chunk recalled. The NI group was not included, since for this condition the precise nature of the recall units was unknown. Only the first entry into successive chunks for instructed conditions was considered. Reentries into previously accessed chunks were rare and accounted for only 2% of the items recalled. A 3 × 4 mixed analysis of variance was conducted on the chunk-size data, with factors representing instructions (I, S, and R) and output position (1–4). No effect of instructions was obtained, since a high level of clustering according to triad membership occurred for all conditions. (The mean chunk size was 2.34 items across conditions, with a standard deviation of .44.) The analysis yielded only a significant main effect of output position, F (3, 126) = 89.10, $p <$ .001, which resulted from a consistent decrease in chunk size across output positions in all treatment groups. This decrease suggests that the recall of items within a unitized grouping may not be solely dependent upon the accessibility of the higher order unit, but rather is subject to qualifications based on position in output. It is conceivable that interference resulting from output processes may be more detrimental to the individual components within a higher order unit than to the units themselves.

Output time. In order to examine the cohesiveness and search rates for higher order units during output, IRTs were obtained from tape recordings of oral recall. The IRTs from *Ss* in the NI condition were not included as they were not relevant to these issues. Interresponse times involving reentries into previously recalled higher order units and extralist intrusions were excluded from the data and represented only 11.2% of the total responses.

A 3 × 2 × 3 mixed analysis of variance was conducted on the mean IRTs for the instructed groups with factors corresponding to instructions (I, S, and R), type of IRT (within or between chunk), and output position (2–4). Only the last 3 output positions were included in this analysis, since there is no satisfactory way to measure access time to the first word of the first unit recalled. The IRT analysis yielded reliable main effects for instructions, F (2, 39) = 5.67, $p < .001$, type of IRT, F (1, 39) = 91.36, $p < .001$, and output position, F (2, 27) = 22.57, $p < .001$. The highly significant difference for between- and within-chunk IRTs indicates that stable higher order units were established

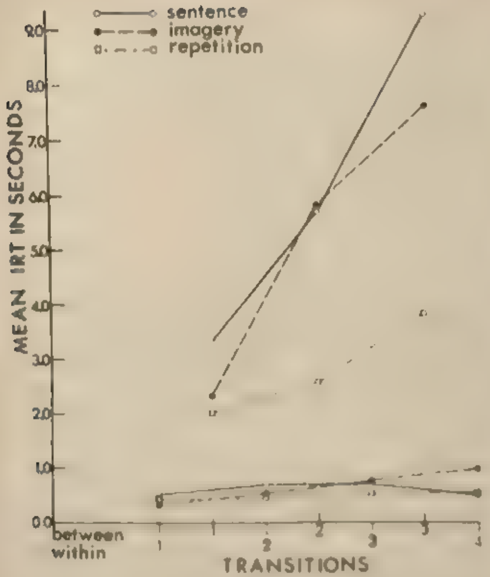


FIG. 2. Mean within- and between-group interword response times (IRTs) as a function of output position for the instructed conditions.

in all instructed groups. The predicted Type of IRT × Output Position interaction was also highly significant, $F(2, 78) = 18.97$, $p < .001$. The between-chunk times can therefore be viewed as representative of the search and retrieval processes for specific higher order units, while the within-chunk times reflect the output of nominal items within higher order units. It is apparent in Figure 2 that the within-chunk IRTs are unaffected by either instructions or output position, suggesting that all higher order units were equally cohesive. The effects of instructions are restricted to the between-group IRTs, as evidenced by the Instructions × Type of IRT interaction, $F(2, 39) = 5.36$, $p < .001$. While the between-chunk times all increased linearly across output positions, instructions had substantial effects on the slopes of these functions and, presumably, the search times for higher order units.

DISCUSSION

The patternings of input and output times strongly indicate that higher order units were formed during encoding by all instructed groups and that these units re[illegible]tional during retrieval. Wherea[illegible] correct responding did not differ am[illegible] ment groups, the input-time data ind[illegible] Ss who organized list materials [illegible] and sentence structures required less [illegible] time to perform the encoding oper[illegible] follows that when Ss are limited [illegible] time constraints, experimental condi[illegible] essarily favor encoding strategie[illegible] image and sentence formation.

One goal of the present research [illegible] to examine the nature of retrieval proce[illegible] higher order units formed intraexperim[illegible] In general, the IRTs observed were con[illegible] with those reported by Pollio et al. (196[illegible] Patterson et al. (1971) with categorized [illegible]ts. However, a major difference is represented by the shape of the between-group functi[illegible] across output positions. Data from the [illegible] which [illegible] paced [illegible] presentation was employed, were described by an exponential function derived from a probabilistic retrieval model. The present results are more consistent with a linear increasi[illegible] [illegible] to that reported by Kel[illegible] et al. (in press). In both the Kellas et [illegible] study and the present investigation, item [illegible]sentation was S-paced, and the high level of correct responding indicated that all u[illegible]ts were highly accessible during the search ph[illegible]se of retrieval. A linear function across output positions could be predicted by at lea[illegible] 2 alternative models of retrieval processes. One is the fixed-search model described, but not empirically supported, by Pollio et al. A linear function could also be predicted by a model based on the trace strength of memory units, as determined by systematic patternings of rehearsal during input (Kellas et al., in press). Rundus (1971) has presented data which indicate that items afforded the most rehearsal during input are typically recalled early in the output sequence. Thus, while the trace-strength model would also assume a fixed retrieval sequence for higher order units, the increase in retrieval times for units at later positions would not result from a reaccess of the memory set at a fixed position, but from an increase in group access times for later units, based on their decreasing trace strengths.

Both linear and exponential functions could be accommodated by a model that includes a parameter related to the accessibility of higher order units. The between-group IRT function would be linear to the extent that units

[illegible] late in the output sequence are ac[illegible] and a consistent retrieval plan remains [illegible]. However, if recall breaks down at [illegible] positions due to inaccessibility of higher [illegible] units, *S* may perform an extended search [illegible] retrieve additional units. This search may [illegible], for example, a repetitive generation [illegible] previously recalled units in an attempt to [illegible] an associative link to unrecalled items, [illegible] by a recognition check for list membership. Consequently, the recall times for [illegible] later units could be expected to increase [illegible]tially.

[illegible] the between-group IRT functions in [illegible] present study are all linear, differences are [illegible] in the slopes of the functions. The [illegible] given repetition instructions were able to [illegible] higher order units significantly faster [illegible] positions than either I or S *Ss*. The [illegible] presented by Patterson et al. (1971) [illegible] that the observed retrieval facilitation [illegible] R condition might best be localized in the group access component. An examination [illegible] suggests an explanation for this facilitation. Little if any increase in processing time [illegible] noted across input triads in either the S or I conditions, while time spent per triad increases nearly linearly in the R condition. This increase provides suggestive evidence that [illegible] rehearsing the earlier presented items [illegible] positions, thereby establishing associa[illegible] among higher order units via cumula[illegible] rehearsal of previously formed triads. This particular encoding process would lead to a decrease in the time required to access additional unrecalled units during recall.

Evidence for the establishment of sequential dependencies in the R condition was obtained from the questionnaires given to all *Ss* following the experimental session. The data from these questionnaires were not previously reported since, for the most part, they simply verify a conclusion drawn from the input-time [illegible] were indeed following instruc[illegible] [illegible] the degree to which *Ss* reported use of cumulative rehearsal in addition to the particular strategy given them, significantly more cumulative rehearsal was reported by *Ss* in [illegible] groups. While such a sequential strategy is [illegible] transitions between units during retrieval.

The high similarity between the IRT data from the present study and research with [illegible] press), suggests a communality in retrieval operations. Pollio et al. (1969) have proposed that retrieval of categorized material is best conceived as a "reconstruction" process, based on the attribute overlap of nominal items to their superordinate label. However, it is difficult to envision such a process operating on higher order units formed intraexperimentally. It seems more parsimonious in both cases to view the retrieval of higher and lower order units as a function of trace strength, determined by the systematic execution of rehearsal processing.

REFERENCES

[illegible], & WINZENZ, D. Group structure, [illegible] memory for digit series. *Journal of* [illegible]

KELLAS, G., ASHCRAFT, M. H., JOHNSON, N. S., & [illegible] *of Verbal Learning and Verbal Behavior*, in press.

[illegible] and type of material on [illegible] term memory. *Journal of Experimental Psychology*, 1971, 88, 50–56.

LAUGHERY, K. R., & SCHMIDT, R. A. Subject pacing and the analysis of the factors in short-term memory. Paper presented at the meeting [illegible] Experimental Psychology [illegible] England, April 1970.

[illegible] In K. [illegible] Spence (Eds.), *The psychology* [illegible] Vol. 1. New York [illegible]

PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971.

[illegible] imagery, and [illegible] 925 nouns. *Journal of Experimental Psychology* [illegible]

PATTERSON, K. E., MELTZER, R. H., & MANDLER, G. Inter-response times in categorized free recall. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 417–426.

POLLIO, H. R., RICHARDS, S., & LUCAS, R. Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 529–536.

RUNDUS, D. Analysis of rehearsal processes in free [illegible] 89, 63 [illegible]

THORNDIKE, E. L., & LORGE, I. [illegible]

(Received April 30, 1973.)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 266-276

STIMULUS SUFFIX EFFECTS IN DICHOTIC MEMORY

STANLEY R. PARKINSON[2] AND LORA L. HUBBARD

Arizona State University

A very-short-term "preperceptual" or "precategorical" auditory buffer has been proposed by Crowder and Morton as the storage system mediating the retention of items in the terminal serial positions of long lists presented sequentially, and by Broadbent as the storage system mediating retention on the ear reported second or "unattended" ear in dichotic tasks. Previous research has shown that terminal serial position performance in long lists presented monaurally or binaurally is selectively impaired when an extra locution (stimulus suffix) is placed at the end of a list. In the present dichotic memory research, the addition of either a monaural stimulus suffix on the unattended ear or a binaural suffix was shown to selectively impair unattended-ear performance. The comparability of stimulus suffix effects in these tasks suggests that the storage systems mediating retention in dichotic memory are identical with the storage systems operating in the retention of long lists of sequentially presented items.

Crowder and Morton (1969) have recently proposed and extensively investigated (e.g., Morton, Crowder, & Prussin, 1971) a model of auditory–verbal information processing. The major emphasis in their empirical work has been to demonstrate the existence of a very-short-term auditory buffer in which information is thought to reside prior to the process of categorization. The stimuli employed in their research have consisted of long lists of verbal items presented monaurally or binaurally, and the dependent variable has been the retention function measured immediately following stimulus presentation.

Their model is based on the assumption that the primacy and recency effects, obtained in serial position analyses of immediate recall functions, are mediated by two distinct storage systems. They maintain that the first few items presented are categorized and are entered into a short-term or active store while the terminal items are held in an ephemeral (1–2 sec.) precategorical acoustic store (PAS).

[1] This research was supported in part by the U.S. Air Force Human Resources Laboratory, monitored by the Flying Training Division under Contract F41609-72-C-0037. The authors wish to thank Mark Knight and Alan Perey for their assistance in running *S*s in these experiments.

[2] Requests for reprints should be sent to Stanley R. Parkinson, Department of Psychology, Arizona State University, Tempe, Arizona 85281.

The acoustic quality of PAS is inferred from data showing that (*a*) the recency effect in serial position functions is generally greater with auditory than with visual presentation (e.g., Corballis, 1966; Crowder & Morton, 1969); (*b*) recency effects of the magnitude of auditory presentation are obtained with visual presentation if Ss vocalize the stimuli during presentation (e.g., Routh, 1970); (*c*) an additional acoustic element placed at the end of a stimulus list (stimulus suffix) selectively impairs performance over the terminal serial positions (Crowder & Morton, 1969; Morton et al., 1971); and (*d*) the stimulus suffix is effective only if both stimulus list and suffix are presented aurally (Morton & Holloway, 1970).

The precategorical nature of PAS is inferred from findings indicating that the magnitude of the stimulus suffix effect is unaffected by either the frequency of occurrence or emotionality of words employed in the suffix position, or the semantic similarity between stimulus list and suffix (Morton et al., 1971).

The model and theoretical constructs proposed by Crowder and Morton (1969) are very similar to those of Broadbent (1971). Working with a dichotic paradigm, Broadbent found that when *S*s were given 2 lists of digits simultaneously, one to each ear, they were more accurate if they first recalled the items from one ear and then

the items from the other (ear-by-ear report) than if they reported items by pair (pair-by-pair report) according to their order of arrival. Broadbent took this finding to imply the existence of a stage of parallel processing (S or storage system) followed by a stage of sequential processing (P or perceptual system). According to the model, both messages of a dichotic pair enter the S system in parallel. A selective filter, operating in time between the S and P systems, selects one message for entry into the P system and attenuates processing of the other message. The P system processes the items in the selected message and that message is reported first during recall. The initially unselected message is held in the S system and that message does not gain access to the perceptual system until the completion of first-ear processing. Like PAS, the level of information processing in the S system is preperceptual or precategorical.

Comparison of the Broadbent (1971) and Crowder and Morton (1969) models reveals that both distinguish between storage systems in which information is relatively unanalyzed, preperceptual or precategorical (S system and PAS, respectively), and storage systems in which information is processed or categorized (P system and short-term or active store). Although the theoretical constructs S system and PAS are derived from different experimental paradigms and hence remain operationally distinct, much theoretical power could be gained by establishing a "working equivalence" between them.

If the storage systems mediating retention in dichotic tasks are the same as those mediating retention with a single auditory channel, we should be able to accurately predict dichotic performance given the single-channel performance. Parkinson (1974) assessed conventional single-channel digit spans and grouped *S*s according to that measure in a dichotic memory task. He found a correlation of +.891 between digit span and overall dichotic performance. This relationship lends considerable support to the merit of establishing a working equivalence between constructs derived from the dichotic task and those derived from a task in which long lists of items are presented over a single auditory channel.

Another way of evaluating a working equivalence between constructs would be to show an effect of an independent variable in one task that is known to be effective in the other task. Precategorical acoustic storage is thought to mediate retention of items occurring in the terminal serial positions of long lists presented monaurally or binaurally. Crowder and Morton (1969) and Morton et al. (1971) have shown that retention of terminal serial position items is selectively impaired by an extra acoustic locution (e.g., "uh") placed at the end of a stimulus list (stimulus suffix). The S system of Broadbent (1971) is thought to mediate the retention of the "unattended" ear or the ear reported second in dichotic tasks. If the S system were equivalent to PAS, a stimulus suffix placed at the end of a dichotic list should selectively impair unattended-ear performance. It is the purpose of the present experiments to explore stimulus suffix effects in dichotic memory.

Experiment I

Method

Subjects and design. Sixteen introductory psychology students served as *S*s in the present experiment. The *S*s were tested individually in a sound reduction chamber and they were paid for their participation. Morton et al. (1971) found that the magnitude of the stimulus suffix effect varied as a function of the similarity of "apparent spatial location" between suffix and stimulus list. Therefore, suffix location was manipulated in the present experiment as a between-*S*s factor. Individual single-channel digit spans were assessed prior to the experimental sessions (Parkinson, 1974), and each *S* was then assigned to one of 4 groups that determined the location of the stimulus suffix to be received: control—no suffix; monaural suffix—attended ear; monaural suffix—unattended ear; and binaural suffix. The method for assigning *S*s to groups roughly equated all 4 groups for mean digit span.

Materials and apparatus. On each trial *S*s received 2 lists of digits simultaneously, one to each ear. All lists were recorded by the same male speaker at the rate of 2 pairs per second. Stimulus lists were recorded on a TEAC model TCA-42 4-channel tape deck and they were played to *S*s via stereo headphones (Superex Pro-BV). The stimulus lists were constructed from the set of

numbers 0–10 and no number appeared twice on a given trial.

The stimuli were recorded on 5 audiotapes, one practice and 4 test. The practice tape was comprised of 42 dichotic lists, each list consisting of 4 digits per ear (8 digits per trial). The 4 test tapes each contained 72 dichotic lists with 5 digits per ear (10 digits per trial). One of these test tapes was used in the first experimental session to establish baseline performance for all 16 *S*s in the dichotic task. In the remaining 3 test tapes a stimulus suffix was placed at the end of each stimulus list. A monaural stimulus suffix was placed at the end of each list in 2 of these tapes, on the right channel for Tape 3 and on the left channel for Tape 4. On Tape 5, a binaural suffix was placed at the end of each stimulus list. In all cases (whether monaural or binaural) the stimulus suffix was the locution "uh." The suffix was recorded in the same voice and in the same tempo as the digit list.

Procedure. According to the Broadbent (1971) model, *S*s attend to one ear during stimulus presentation and report that ear first during recall. In previous research (Bryden, 1971; Parkinson, 1974), the ear of attention has been controlled by instructing each *S* to "attend" to a particular ear during the course of an experiment. A payoff schedule was introduced in the present experiments to meet the same purpose. Each *S* was told that he would receive 1 cent for each digit recalled in the correct order for one ear, hereafter referred to as attended ear, and ½ cent for each digit on the other ear, hereafter called unattended ear. The ear receiving maximum payment was balanced within each group (2 left ear and 2 right ear) and, for each *S*, remained constant during the course of the experiment. The *S*s were told to earn as much money as they could, and feedback regarding the amount earned was given following *S*'s response on each trial.

The *S*s were instructed to report ear by ear. Following a procedure developed by Bryden (1971), order of report was manipulated, i.e., on one half of the trials *S*s were instructed to report attended ear first and unattended ear second (A–U report), and on the other half of the trials unattended ear first and attended ear second (U–A report). Unlike the Bryden procedure, in which order of report was blocked, report instructions were delayed until the end of each stimulus list and the order of report was randomized. The ear to be recalled first (left or right; for *S*, attended or unattended) was indicated by the onset of one of 2 lights at the end of each stimulus list. The lights, one red and one green, were oriented horizontally on a panel in front of *S*s to coincide with right and left ears. Tone bursts recorded on a third channel of the tape triggered the lights to come on immediately following list presentation. In the suffix conditions, the indicator light was presented immediately following the suffix. It was thought that the delayed order-of-report instructions coupled with the payoff schedule would encourage *S*s to differentially process attended- and unattended-ear messages.

At no time were *S*s required to recall all 10 items. The *S*s were instructed to report only those items they actually remembered from the unattended ear. In order to receive maximum payment for his report on each trial, *S* had to follow order-of-report instructions as indicated by the lights, i.e., report digits in correct serial order for each [illegible] The *S*s' verbal responses were written down by E and compared with mimeographed copies of [illegible] dichotic lists actually recorded on each of the [illegible] Following *S*'s report on each trial, *E* assessed errors made by *S* and told him how much he earned for that trial. Although serial recall scoring was used in determining payment, free recall scoring was used in data analysis.

The *S*s from all 3 suffix groups were told before the second test session that an extra locution ("uh") would occur following each list, and they were informed which ear would be receiving the suffix. They were told to continue to do what they had been doing in the previous session, to ignore the suffix, and not to report it during recall. In the second test session the control group repeated the session with the first test tape with headphones reversed so that attended and unattended messages were reversed.

Results

Raw scores were converted to percentage correct recall and separate 2 × 3 (Between × Within) analyses of variance (ANOVAs) were conducted on data obtained in the first and second test sessions. Between-groups factors in these analyses were ear of attention and suffix group. Within-groups factors were order of report (A–U vs. U–A), attention (A vs. U), and serial position. Baseline and suffix performance is shown in Figures 1–4 for the control group, attended-ear suffix group, unattended-ear suffix group, and binaural suffix group, respectively. As there was no interaction between order of report and type of suffix, performance in both sessions is shown collapsed over order of report.

Baseline test. Comparison of the baseline performance shown in Figures 1–4 indicates that prior to experimental treatment (administration of a stimulus suffix) the 4 groups were not different. This was confirmed in statistical analysis, as neither the main suffix group factor nor any of the suffix group interaction factors were reliable, all $ps > .05$.

As in previous research in which order of report was manipulated (Bryden, 1971;

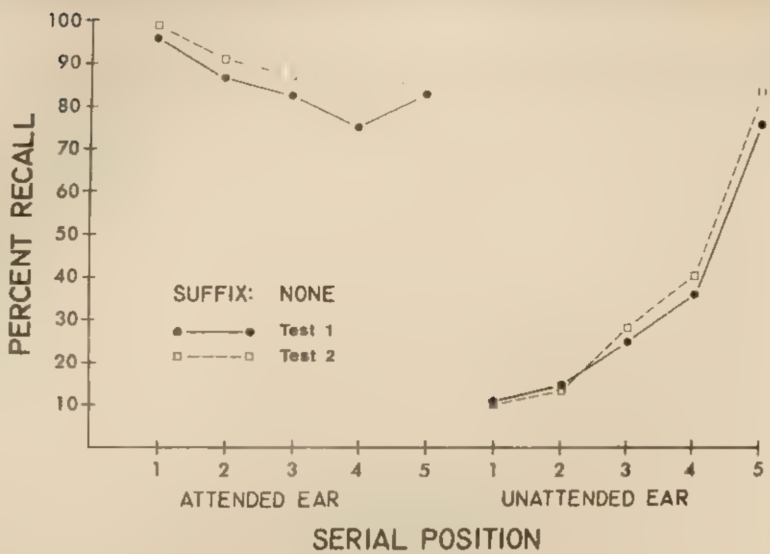


FIGURE 1. Control group, Experiment I: Percentage recall in the first and second test sessions as a function of ear of attention and serial position.

Parkinson, 1974), performance on the attended ear was superior to performance on the unattended ear, $F\ (1, 8) = 191.31$, $p < .001$, and A–U reports were superior to U–A reports, $F\ (1, 8) = 17.64$, $p < .003$. Serial position functions for the attended ear were relatively flat while positively increasing serial position functions were obtained for the unattended ear, $F\ (4, 32) = 101.52$, $p < .001$.

Suffix test. Based on the Broadbent (1971) model and on the findings of Morton et al. (1971), we would predict that stimulus suffix effects would be limited to the terminal serial positions of the unattended ear in conditions in which the un-

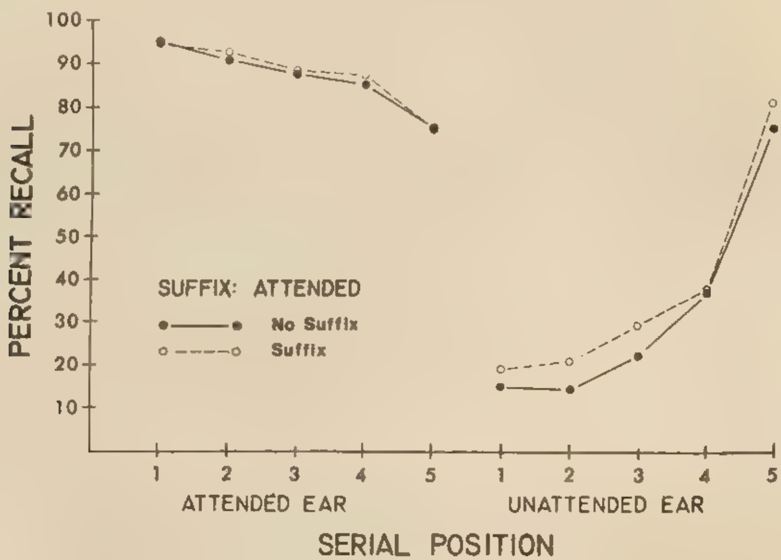


FIGURE 2. Attended-ear suffix group, Experiment I: Percentage correct recall as a function of test (suffix vs. no suffix), ear of attention, and serial position.

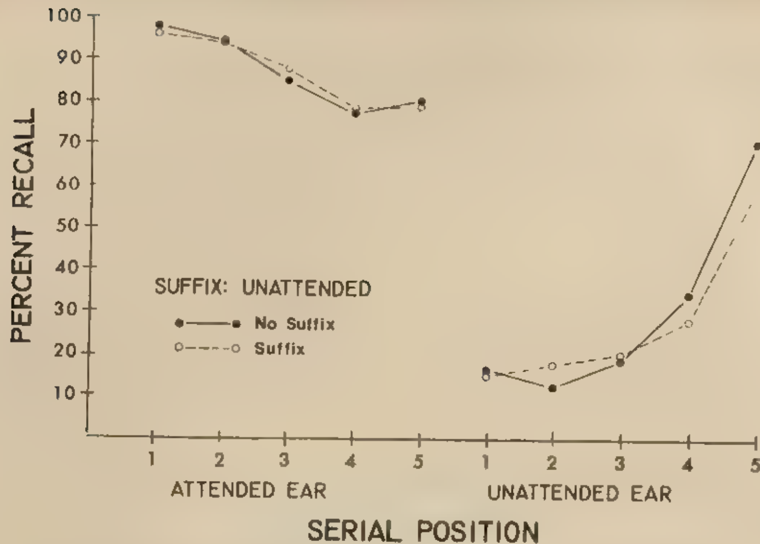


FIGURE 3. Unattended-ear suffix group, Experiment I: Percentage correct recall as a function of test (suffix vs. no suffix), ear of attention, and serial position.

attended ear received input subsequent to the stimulus list. In the present experiment this occurred in only the unattended suffix and binaural suffix groups. No effect should be observed in the attended suffix group.

Inspection of Figures 1–4 reveals that stimulus suffix effects were indeed limited to the terminal serial position of the un attended ear in the unattended suffix an binaural suffix groups. Results of t ANOVA indicated a reliable Suffix Gro × Serial Position interaction, F (12, 32 2.41, $p < .02$, and a Suffix Group × tention × Serial Position interaction. (12, 32) = 2.64, $p < .01$. Neither the

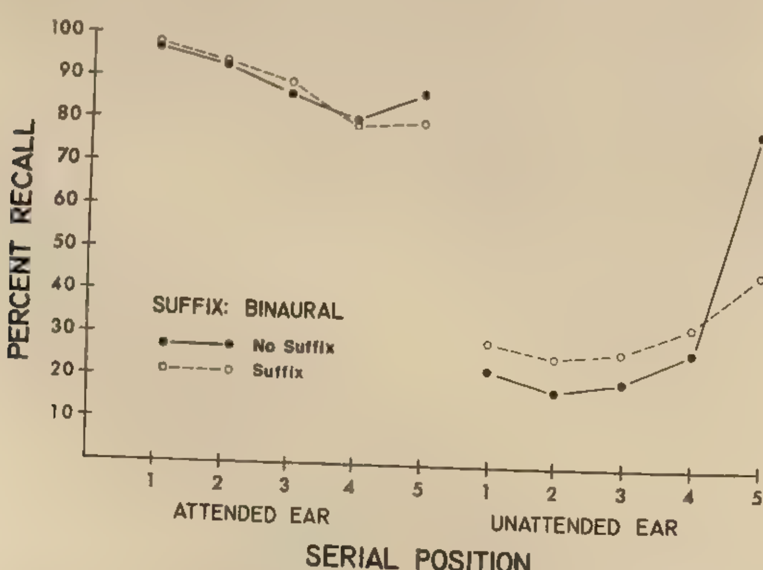


FIGURE 4. Binaural suffix group, Experiment I: Percentage correct recall as a function of test (suffix vs. no suffix), ear of attention, and serial position.

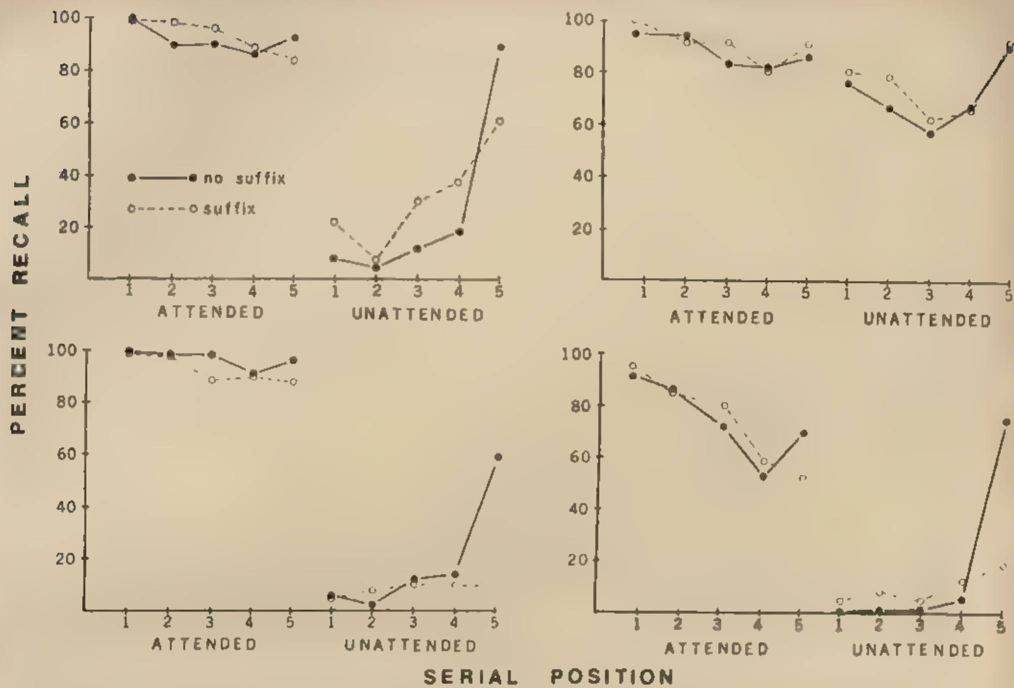


FIGURE 5. Individual data ($N=4$), binaural suffix group, Experiment I: Percentage correct recall as a function of test (suffix vs. no suffix), ear of attention, and serial position.

main effect of suffix group nor any of the other suffix group interactions were reliable.

Difference scores between baseline and suffix conditions were obtained for all *Ss* on the terminal serial position of the unattended ear. As a test of the reliability of the suffix effects, *t* tests were conducted on the difference scores. The *t* tests showed reliable decrements in the unattended suffix group, $t\ (6) = 3.15$, $p < .02$, and the binaural suffix group, $t\ (6) = 3.44$, $p < .02$, relative to the control group, but no difference between the control and the attended suffix groups, $t\ (6) < 1$.

The comparability of suffix effects in the present experiment with those in the Morton et al. (1971) study suggests that the storage systems mediating retention in dichotic memory are identical with the storage systems operating in the retention of long lists of sequentially presented items.

The binaural suffix group had the largest stimulus suffix effect (mean decrement on the terminal serial position = 33.25%); however, there was considerable variability between *Ss* in this group with regard to the magnitude of the suffix effect. Considering only the terminal serial position, the suffix effects of the 4 *Ss* in the binaural suffix group were 55%, 50%, 28%, and 0 (see Figure 5). Of considerable interest is the fact that there was an orderly relation between magnitude of suffix effect and digit span. The *Ss* showing the largest suffix effects had digit spans of 5 and 7, the *S* with an intermediate suffix effect had a digit span of 9, and the *S* with no demonstrable suffix effect had a digit span of 11. Experiment II was conducted in order to further delineate the relation between magnitude of suffix effect and digit span.

EXPERIMENT II

Method

Digit spans were assessed and only individuals with a high or low digit span were selected for participation as *Ss*. Sixteen *Ss* were assigned on

the basis of digit span to one of 2 groups: high span or low span. The Ss in the high-span group had individual digit spans of 10, 10, 9, 10, 9, 10, 10, and 10, with a mean of 9.75. Ss in the low-span group had individual spans of 6, 7, 7, 6, 6, 6, 6, and 6, with a mean of 6.25. The Ss received the same test sequence (one practice session followed by 2 test sessions) and same instructions as those employed in the first experiment. One half of the Ss in each group received maximum payment for right-ear performance and one half for left-ear performance.

In the present experiment, the stimulus suffix ("uh") was presented binaurally for both groups.

Results

As the primary concern in Experiment II was the relation between digit span and stimulus suffix effect magnitude, the data were collapsed over order of report and separate 2 × 2 (Between × Within) ANOVAs were conducted for attended-ear reports and for unattended-ear reports. In these ANOVAs, span and ear of attention were between-groups factors while test (baseline vs. suffix) and serial position were within-groups factors. The results for the low-span group and for the [illegible]

low-span group to be a positively increas[illegible] function of serial position. The [illegible] responding function of the high-span gr[illegible] (Figure 7) shows level of accuracy to [illegible] relatively flat over the first 4 serial [illegible] tions, with somewhat higher perform[illegible] on the terminal item. Statistical ana[illegible] revealed significant main effects of [illegible] F (1, 12) = 157.25, $p < .001$, and serial position, F (4, 48) = 22.84, $p < .001$, and a reliable Span × Serial Position interaction, F (4, 48) = 3.01, $p < .03$. These findings provide confirmation of earlier results (Parkinson, 1974) showing an [illegible] relation between unattended-ear rec[illegible] curacy and digit span.

Inspection of Figures 6 and 7 shows [illegible] the effect of a binaural stimulus suffix [illegible] limited to the terminal serial position, and statistical analysis revealed the Test (b[illegible] line vs. suffix) × Serial Position int[illegible] action to be reliable, F (4, 48) = 2[illegible] $p < .001$. This result replicates [illegible] obtained with the binaural suffix group [illegible] Experiment I and provides a clear dem[illegible] stration of a stimulus suffix effect in dicho[illegible] memory. Of further interest is the mag[illegible] tude of the stimulus suffix effect. The mean decrement in the terminal serial position (baseline suffix) for the low-span group

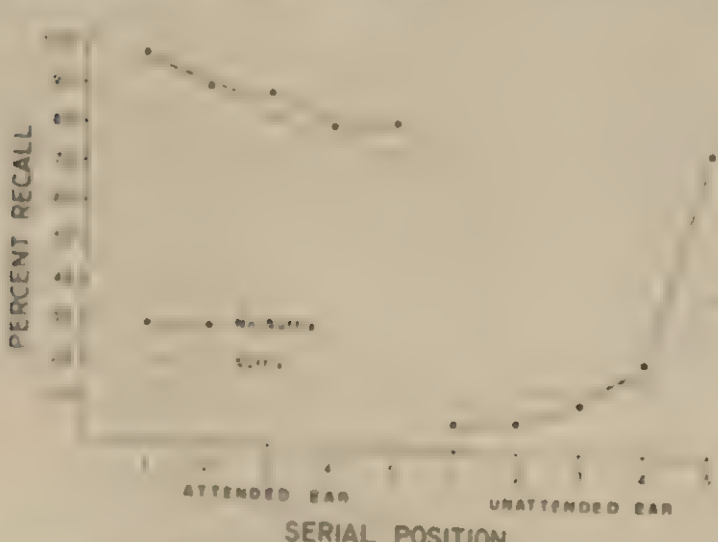


FIGURE 6. Low-span group, Experiment II. Percentage correct recall as [illegible]

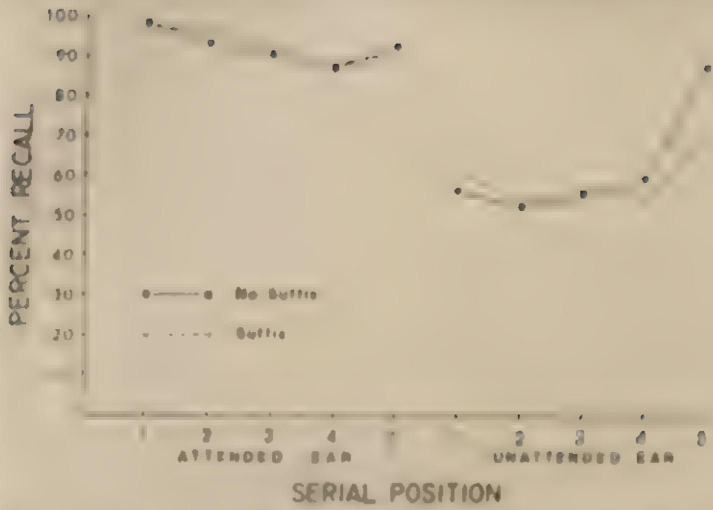


FIGURE 2. High span group, Experiment II. Percentage correct recall as a function of test (suffix vs. no suffix), ear of attention, and serial position.

[illegible]%, while that for the high span [illegible] 16.04%. The Span × Test × Serial Position interaction was reliable, F (4, 48) = 3.01, $p < .05$.

[illegible] ear recall. The effect of span [illegible] limited to the unattended ear [illegible] span group also performed more [illegible] than the low span group on the [illegible], F (1, 12) = 12.16, $p < .005$, with the greatest disparity between groups on the terminal serial positions, F (4, 48) = [illegible], $p <$ [illegible].

[illegible] was also an effect of the stimulus [illegible] on the attended ear, as indicated by [illegible] × Serial Position interaction, F [illegible] control to the low span group, F (1, 12) = [illegible], $p < .05$. The third-order interaction between span, test, and serial position was not reliable.

Discussion

[illegible] group of Experiment I is now clear, [illegible] stimulus suffix effects [illegible]

[illegible] span [illegible] the nature of the difference between [illegible] high and low digit span which results in different magnitudes of the stimulus suffix effect? The most obvious possibility is that the magnitude of the stimulus suffix [illegible] the amount of stimulus [illegible]

[illegible]

[illegible] unattended ear [illegible] digit correctly recalled [illegible]

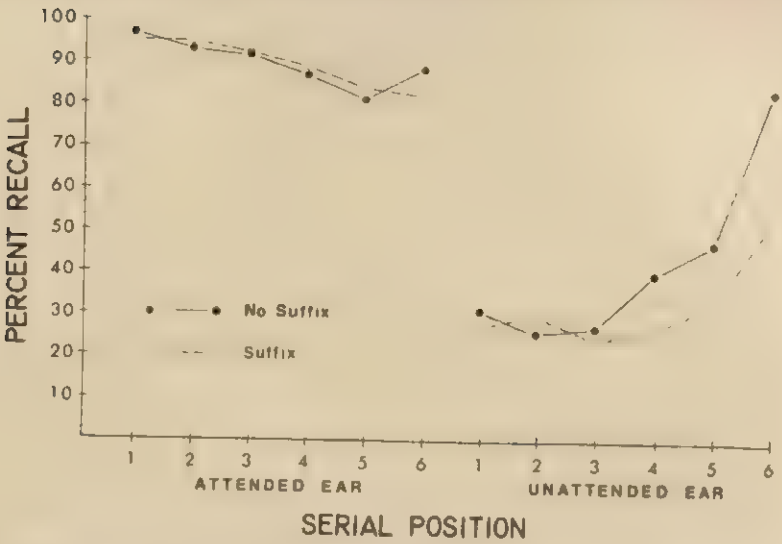


FIGURE 8. Group data, Experiment III: Percentage correct recall as a function of test (suffix vs. no suffix), ear of attention, and serial position.

Results

As in Experiment II, the data were collapsed over order of report and separate ANOVAs were conducted on attended-ear and unattended-ear performance. These results are shown in Figure 8.

Attended-ear recall. The only effect reliable in attended-ear analyses was the main effect of serial position, with primary serial positions better recalled than terminal serial positions, $F(5, 30) = 8.81$, $p < .001$.

Unattended-ear recall. Baseline performance was superior to that observed when a stimulus suffix was administered, $F(1, 6) = 10.01$, $p < .02$, and terminal serial positions were recalled with greater accuracy than primary serial positions, $F(5, 30) = 13.49$, $p < .001$. Inspection of Figure 8 reveals the stimulus suffix effect to be restricted to the terminal serial positions, and statistical analysis showed the Test $\times$ Serial Position interaction to be reliable, $F(5, 30) = 4.90$, $p < .002$.

Of considerable interest is the resemblance of the unattended-ear serial position function for the high-span *S*s in the present experiment (with dichotic lists containing 6 digits per ear) to the unattended-ear serial position function of the low-span *S*s in Experiment II (with dichotic lists containing 5 digits per ear). Both of these show unattended-ear accuracy to be a positively increasing function of serial position. There is also considerable similarity in terms of the magnitude of the stimulus suffix effect in these functions. The mean decrement on the terminal serial position (baseline-suffix) was 28.75% in the present experiment. This can be compared with the 29.16% and 16.04% decrements shown by the low-span and high-span groups, respectively, in Experiment II. A difference score (baseline–suffix) was computed for the terminal serial position performance of high-span *S*s in Experiment II and Experiment III. A correlated *t* test of these differences indicated a reliably larger decrement in Experiment III, $t(7) = 2.51$, $p < .05$.

GENERAL DISCUSSION

According to the Broadbent (1971) model, attention is directed to one message in the dichotic task. That message enters the P or perceptual system while the initially unattended message is held in a short-term auditory buffer, the S system. In this model information is held in different storage systems and these systems can be characterized by different levels of processing. Information held in the S system is relatively unanalyzed while information in the P system is cate-

gorized. Given this type of model we would expect (*a*) differences in the recall functions of attended and unattended messages, and (*b*) differential effects on attended and unattended messages from an independent variable (stimulus suffix) which is thought to selectively interfere with relatively unanalyzed information. Both of these expectations were realized in the present experiments.

Another finding of considerable interest in the present experiments was that the magnitude of the suffix effect was directly proportional to the amount of stimulus overload, i.e., the degree of disparity between an individual's digit span and the number of to-be-remembered items. In Experiment II, given messages containing 5 digits per ear, *S*s with low digit spans showed larger stimulus suffix effects than did *S*s with high digit spans. That it was the disparity between digit span and number of items, and not digit span per se, which was important, was revealed in Experiment III, in which *S*s with high digit spans showed larger stimulus suffix effects when they were given dichotic lists containing 6 digits per ear.

How can we interpret the relationship between digit span and suffix effect? First consider results obtained by Parkinson (1974). He manipulated order of report in dichotic memory and found correlations of +.881 and +.844 between digit span and unattended-ear performance in attended–unattended and unattended–attended reports, respectively. Parkinson offered two interpretations of his results one based on S system storage capacity and one based on "central processing capacity." Attempts to delineate the relationship between digit span and suffix effect magnitude will be oriented along these interpretations.

Storage Capacity

One interpretation of the relation between digit span and unattended-ear performance is that S system capacity varies as a function of digit span. Accordingly, *S*s with high digit spans have more "storage bins" in their S system than do *S*s with lower digit spans. The stimulus suffix in the present experiments was presented in the same voice and tempo as the stimulus list; hence, the relation was one of high similarity. Morton et al. (1971) manipulated similarity between stimulus list and suffix along dimensions of intensity, pitch, and apparent spatial location. They found the magnitude of the stimulus suffix effect to be directly proportional to the degree of similarity between suffix and stimulus list. Reports of *S*s in the present experiments indicated that the degree of similarity between the suffix and the stimulus list made it very difficult for them to ignore the suffix during presentation. One way of viewing this effect is that as the similarity between stimulus list and suffix increases, *S*s find it more difficult to separate out or isolate the suffix from the stimulus list and the suffix comes to "act" like an additional item in the list. Given a system of storage bins and a displacement principle, the suffix would simply displace the last item in the list as the last stimulus item displaces the penultimate item. Given sufficient storage bins (higher-span *S*s), the displacement would not affect recall to any great extent; but given more items than bins (low-span *S*s in Experiment II and high-span *S*s with more items as in Experiment III), the effect of a suffix would be to even further overload the system and the result would be a significant impairment of recall performance.

Central Processing Capacity

In some models of information processing, attention is used synonymously with central processing capacity (CPC) or "mental effort" (e.g., Kahneman, 1973; Posner & Boies, 1971). Models of this type posit a finite amount of CPC that is available for information processing. Mental operations requiring CPC share a common source; thus, if a large amount of CPC is invested in one operation, the amount remaining for others is reduced.

Consider that the present task consists of 2 components, recall and separation or isolation of the suffix from the stimulus list, both of which require CPC. The effects on recall observed following manipulation of similarity between stimulus list and suffix (Morton et al., 1971) can be viewed in terms of the amount of CPC required by the isolation component. In the high-similarity conditions, more CPC was required by the isolation process than in the low-similarity conditions; therefore the amount remaining for recall was insufficient to maintain accurate performance.

In Experiment II of the present research, both groups, high and low span, were given binaural suffix presentation in the same voice and tempo as the stimulus list. We can assume from the findings of Morton et al. (1971) that some amount of CPC was demanded in the present experiment to separate

the stimulus list and suffix. Parkinson (1974) hypothesized that the CPC required for recall varied in direct proportion to the difference between an individual's digit span and the number of to-be-remembered items. Assuming Parkinson's hypothesis to be correct, more CPC was demanded in the recall task of Experiment II for the low-span group than for the high-span group. Given a condition in which substantial CPC is needed for recall, any CPC diverted to the separation of stimulus list and suffix should be sufficient to impair recall performance. We would expect on this basis a greater discrepancy between baseline and suffix test performance for the low-span group than for the high-span group. We would also expect that as CPC requirements for recall were increased for the high-span group by increasing the number of items (Experiment III) a greater difference would exist between baseline and suffix performance. Both of these results were obtained in the present experiments.

The purpose of the present experiments was to evaluate the merit of establishing a working equivalence between theoretical constructs derived from dichotic memory and from procedures in which long lists of items are presented sequentially. The comparability of stimulus suffix effects in the present experiments to those of Morton et al. (1971) suggests that the storage systems mediating retention in dichotic memory are identical with the storage systems operating in the retention of sequentially presented items

REFERENCES

Broadbent, D. E. *De... n and stress.* London: Academic Press, 1971

Bryden, M. P. Atte... trategies and short-term memory in d... stening. *Cognitive Psychology*, 1971, 2, 9...

Corballis, M. C. R... d decay in immediate recall of vi... ally presented items. *Canadian Jou... gy*, 1966, 20, 43-51.

Crowder, R. G., & M... Precategorical acoustic storage (PA... *ion & Psychophysics*, 1969, 5, 365...

Kahneman, D. *Atten...* Englewood Cliffs, N.J.: Prentice-H...

Morton, J., Crowder, R. G., & Prussin, H. A. Experiments with t... ffix effect. *Journal of Experiment... Monograph*, 1971, 91, 169-190.

Morton, J., & Holloway, C... Absence of cross-modal "suffix effe... term memory. *Quarterly Journal of E... Psychology*, 1970, 22, 167-176.

Parkinson, S. R. Variability and control in dichotic memory. *Journal of Experimental Psychology*, 1974, 102, 67-80.

Posner, M. I., & Boies, S. J. Components of attention. *Psychological R...*, 1971, 78, 391-408.

Routh, D. A. "Trace strength," modality, and the serial position curve in immediate memory. *Psychonomic Science*, 1970, 18, 355-357.

(Received May 3, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 277–283

RETROACTIVE INHIBITION OF RHYME CATEGORIES IN FREE RECALL:

INACCESSIBILITY AND UNAVAILABILITY OF INFORMATION[1]

DOUGLAS L. NELSON[2] AND DAVID H. BROOKS

University of South Florida

This experiment was designed to investigate retroactive inhibition in free recall of categorized word lists with categories defined as sets of rhymes. Each *S* learned 1–6 lists prior to being tested for total free recall of all words from all lists. Following this test, half of the *S*s were given a second total free-recall test and half of the *S*s were given a cued-recall test. The cues consisted of sequences of letters which, when pronounced, rhymed with 4 of the words from the list. The results indicated that retroaction effects were evident for the recall of words, categories, and for words within recalled categories. Providing retrieval cues eliminated retroaction effects but, simultaneously, depressed the number of words recalled per category. An analysis of extralist intrusions suggested that the loss of instances within recalled categories arose because of the difficulty of discriminating between intralist and extralist instances.

Tulving and Psotka (1971) have shown that retroactive inhibition in the free recall of categorized word lists is apparently produced by the loss of higher order memory units rather than instances falling within those units. Furthermore, since providing category names as retrieval cues essentially eliminated retroaction, they concluded that this loss reflected the inaccessibility rather than the unavailability of the higher order units. These findings were interpreted as indicating that words of a conceptual category are "stored together" as a functional unit. Providing the name for an apparently forgotten memory unit ostensibly provided retrieval access to most of the instances coded within it.

This experiment was concerned with whether or not rhyme categories could serve as functional units of memory and with determining if rhymes could serve as effective retrieval cues for recalling instances coded within these categories. To implement these aims, the basic procedures of the Tulving and Psotka (1971) study were replicated using rhyme-defined categories. Findings were expected to be relevant to recent conceptualizations of word-coding processes. Despite the growing consensus that words are represented in memory as sets of features, attributes, or dimensions (e.g., Gibson, 1971; Underwood, 1969; Wickens, 1970), very little research on organizational processes in free recall on other than semantic attributes has been reported. Thus, one aim of this experiment was to show that the forgetting of rhyme categories could be described as an analogue of the forgetting of conceptual categories, or alternatively, to show that memory processes involved in phonetic representation are different from those involved in semantic representation.

The experiment was designed to provide evidence on 2 specific questions. First, does the acquisition of other lists reduce the recall of higher order units, instances within those units, or both? If rhymes are organized into functional units during the study phase of the task, i.e., if they are in some sense stored together in a manner analogous to conceptual categories, then retrieval of the rhyme category ought to mediate the recall of most of its instances. If this supposition is correct, retroaction ought to be reflected in the

[1] This research was supported by Grant MH-16360 to the first author from the National Institute of Mental Health, and by Stipend GZ2341 from the National Science Foundation to the second author. The authors are indebted to Eileen Edmunson, Valerie Reed, and Barbara Mikelait for collecting the data for this experiment.

[2] Requests for reprints should be sent to Douglas L. Nelson, Department of Psychology, University of South Florida, Tampa, Florida 33620.

recall of categories, not in the recall of instances within categories. However, there are inherent differences between the 2 types of categories, differences which may affect the manner of the processing of instances. If rhyme categories are not represented in memory as units, then retroaction may be apparent only within the number of words recalled from each list. Alternatively, rhyme categories may be represented in memory as subjectively organized units, but because of their unique qualities, recalling the unit or category may be insufficient for mediating recall of its instances. If so, then retroaction effects would be reflected in category recall and in the recall of words within recalled categories.

The second question of interest concerned the status of information loss produced by the learning of other structured lists. Does retroactive inhibition reflect the unavailability of information, its inaccessibility, or both? Evidence relating to this question will be provided by presentation of rhyme categories as retrieval cues following free recall. If retroaction is produced by a loss of information from storage within the context of the task, rhyme cues should be ineffective as retrieval aids. If retroaction is produced only by the inaccessibility of information, these retrieval cues might provide access to the information so that forgotten items might be recovered. Of course, it is possible that these cues may be differentially effective at the levels of "units" and "instances." Providing rhymes as retrieval cues may provide access to forgotten categories, but these categories, in turn, may be insufficient for generating most of the list words occurring within these categories. In this case, retroaction would reflect the inaccessibility of higher order units and the unavailability of instances within those units.

Method

Design. There were 8 independent groups of *S*s differing in terms of number of lists learned, and whether second total recall of all words from all lists was either free or cued. Thus, *S*s learned 1, 3, or 6 different lists; following the first total free-recall test, half of the *S*s were given a second total free-recall test and half were given a [c]ued-recall test. Controls acquired a single list, [fol]lowing a delay appropriate for the first list lear[ned] in the 6 list condi ions and following the first [tota]l free recall, half of the controls were given [a] second total free-recall test and half were give[n a] cued-recall test. The controls will be referred [to] as 1D, i.e., a one-list delay condition. Thus, t[he] experimental design conformed to a 4 × 2 × [2] mixed-model factorial, the between-*S*s facto[rs] being number of lists learned (1D, 1, 3, or 6) [and] type of recall (free or cued), and the within[-*S*s] factor being order of recall (first total recall [vs.] second total recall). Since the first total reca[ll] was always free, comparisons of type of re[call] at this point provide information only on the [com]parability of the free and cued groups. Therefore, given this design, meaningful comparisons will b[e] reflected only in interactions between type of r[ecall] and order of recall. That is, for a given m[easure] of performance the effects of cued recall [wou]ld be apparent on the second test and not on [th]e first test. Type of recall was made a betwee[n-*S*s] variable because of the significant improv[eme]nt in performance from the first "total" free-[reca]ll test to the second "total" free-recall test o[bta]ined by Tulving and Psotka (1971). To the [exten]t that performance improves with successive recall tests, some of the improvement obtained in [cued] recall might have been reasonably attributed [t]o this factor. However, because cued recall was [prec]eded by free recall, it was also possible with this design to evaluate the effects of cuing as a [wit]hin-*S*s variable.

Verbal material. Six different 24-word lists were constructed with each list containing 6 categories of 4 words. Items within each category, all relatively common words, consisted of monosyllabic rhymes taken from Wood's (1971) unabridged *New World Rhyming Dictionary*. Thus, with 6 lists and 6 categories/list, there was a total of 36 different rhyme categories. Within a given category of 4 words, all items were both phonetically and visually similar, sharing 2–4 letters in the terminal positions. According to a tabulation and count of words selected from Wood's dictionary there was an average of 15.00 (SD = 4.88) words sharing each ending for each of the 36 categories used in this experiment. The smallest set contained 8 items and the largest contained 24 items. All first letters within a given rhyme category were different and were consonants. In addition to these constraints, no homophones were used in any of the lists, and an attempt was made to minimize obvious associative relationships among items in all lists. Furthermore, an attempt was made to maximize the number of unique word beginnings. There was a total of 42 different beginnings for the 144 words, counting B . . . as different from BR . . . .

In each list all words of a given rhyme category were presented in immediate succession, in a "block," e.g., DITCH, WITCH, STITCH, PITCH, BOOT, LOOT, ROOT, SHOOT, etc. Latin squares were used

to counterbalance the order of appearance of the categories within a given list as well as the ordinal position of lists for groups acquiring more than a single list. The order of the items within a given rhyme category was unsystematically randomized for each *S*.

Procedure. All *S*s participated in individual sessions. In immediate recall, the words were shown, one at a time, by a Kodak Carousel slide projector driven by an auxiliary timer so that each word was shown for 1 sec. Every list was shown 3 times in succession, each time in a different order, before its recall was tested. There was a 10-sec. interval separating successive inputs. Following the third presentation, *S*s were allowed 90 sec. to write down as many words as could be recalled in any order. Guessing was encouraged but not required. The recall sheet contained 24 unnumbered, horizontal lines arranged into 6 groups of 4 lines. Upon completion of the recall phase *S* handed his protocol to *E* who provided a new sheet if there was another list. The next list was presented as soon as the recall sheets were exchanged.

The immediate recall test of the last list was followed by 5 min. of interpolated activity consisting of solving multiplication and long division problems. Speed and accuracy were emphasized. Since the 6-list condition required 25 min. for the immediate recall series, *S*s in the 1D condition engaged in this task for 30 min. so that this interval was equivalent to 5 lists plus 5 min. of interpolation.

Following the interpolated task, all *S*s were given a total free-recall test. Each *S* was provided with new recall sheets and was instructed to write down as many words from all previous lists as he could, in any order. Recall time was equal to 90 sec. multiplied by the number of lists. After the protocols were collected, half of the *S*s in each condition were asked to provide a second total free recall and the other half of the *S*s in each condition were given instructions for cued recall. The latter indicated that all words from all lists were to be recalled but that cues would be provided on the recall sheets to aid them. They were informed that these cues would consist of groups of letters which, when pronounced, would rhyme with 4 of the words presented, e.g., "$\overline{\text{OOT}}$" for BOOT, LOOT, etc. Each cued-recall sheet corresponded to one list arranged so that each rhyme cue was followed by 4 lines for recording the responses. The order of the category names for a given list was randomized and was unrelated to order of appearance during input, but the order of the lists corresponded to that used at input.

Subjects. Eighteen *S*s were assigned to each of the 8 between-*S*s treatments, making a total of 144 *S*s. These *S*s were assigned in blocks of 8, with 1 *S* from each condition per block. Assignment within blocks was determined by a table of random numbers. All *S*s were selected from introductory psychology courses and received points toward their grades for participation. Each of 3 female *E*s ran one-third of the *S*s in each treatment so that *E* was balanced with respect to all conditions.

Results

Separate analyses of variance were performed on immediate recall within each list condition, including a separate breakdown for type of recall. All of these analyses showed that type of recall was insignificant, attesting to the comparability of the free- and cued-recall groups on immediate recall. The mean number of words recalled per list, with the data pooled across all lists in the 8 groups, was 13.86. All deviations from this value, for all lists and conditions of recall, were less than one word. Similar analyses of variance were performed on categories and on words/category. For both of these measures, all analyses indicated that type of recall was insignificant and that recall per list was constant. The mean number of categories per list represented in immediate recall was 4.95, and the mean number of words/category was 2.80.

The results of a clustering analysis of the immediate recall data were consistent with the supposition that rhyme categories can serve as functional units of memory. Ratios of number of items recalled in clusters to total recall were calculated with a cluster defined as at least 2 adjacently recalled rhyming words (Marshall, 1967). These values indicated that nearly all of the words were recalled in rhyme groupings, and that this tendency was not affected by amount of practice. For example, for *S*s learning 6 lists, 89% of the words recalled from the first list were recalled in categories and 86% of the words recalled from the sixth list were recalled in categories. To the extent that *S*s attempted to recall the list in terms of order of input, these values would tend to overestimate the degree of clustering. Nevertheless, each list was presented in 3 different input orders prior to immediate recall, a procedure that should have discouraged the utilization of a seriation strategy.

One of the principal findings of this experiment was that the words/category measure, in addition to the word and

TABLE 1

Mean First-List Recall of Words, Categories, and Words/Category for the First Total Free-Recall Test

Measure	Number of lists			
	1D[a]	1	3	6
Words	11.00	12.58	6.36	4.50
Categories	4.11	4.56	2.53	1.80
Words/category	2.56	2.73	2.55	2.11

[a] The abbreviation 1D = one-list delay condition presented to control *Ss*.

category recall measures, varied as a function of the number of interpolated lists. Table 1 displays mean recall of words, categories, and words/category from the first list learned as a function of number of acquired lists. These data were obtained from the first total free-recall test. Separate analyses of variance indicated that number of lists learned was a reliable source of variance for words, F (3, 140) = 34.86, $p < .01$, categories, F (3, 140) = 40.37, $p < .01$, and words/category, F (3, 140) = 3.93, $p < .05$. Fisher's least significant difference (LSD) values for each of these respective measures were 1.79, .57, and .38. Thus, the results of these analyses confirm the obvious trends displayed in the table. Significant and increasing amounts of retroaction on the initial list were obtained, for every performance measure, as the number of interpolated lists was increased from 0 to 2 to 5. [illegible]ther-more, performance on 1D fell betw[illegible] that obtained for the 1- and 3-list co[illegible]ions. However, even though the passag[illegible] time resulted in some apparent forge[illegible], the magnitude of this loss was small [illegible]lative to that produced by retroaction.

Table 2 presents the mean n[illegible]r of words recalled per list on the fir[illegible]d on the second total recall tests. [illegible]arate analyses were performed for each [illegible] condition with type of recall as a betw[illegible]en-*Ss* factor and with order of recall an[illegible] when appropriate, with list position a[illegible]ated measurements variables. Since [illegible]nclusions reached by within-*Ss* com[illegible]sons were identical to those reached [illegible] between-*Ss* comparisons, only the lat[illegible] will be discussed. These analyses all r[illegible]vealed significant Type of Recall × Order [illegible] Recall interactions: For the 1D, 1, [illegible] and 6-list conditions, the obtained Fs [illegible] this interaction were, respectively, F [illegible] 34) = 12.83, 8.23, 29.56, and 63.10, p [illegible] .01. Each of these interactions indica[illegible] that the facilitating effects of cuing w[illegible] obtained for only the second total recall test, and hence, these analyses attested to the comparability of the free- and cued recall groups on the first total recall test. The sizable amounts of retroactive inhibition obtained in first total recall for both free and cued groups and in second total recall for the free group were completely elimi-

TABLE 2

Mean Number of Words Recalled on the First and Second Total Recall Tests

[illegible] and type of recall	1	2	3	4	5	6	M	[illegible] 1	2	3	4	5	6	M
1D[a]														
Free	11.22						11.22	11.50						11.50
Cued	10.78						10.78	13.40						13.40
1														
Free	12.50						12.50	13.00						13.00
Cued	12.67						12.67	15.56						15.56
3														
Free	[illegible]	9.17	12.33				9.57	8.44	10.11	13.06				10.54
Cued	5.50	8.17	10.33				8.00	11.94	12.39	12.61				12.31
6														
Free	4.61	6.22	7.50	8.67	8.56	10.83	7.73	4.89	6.98	7.11	8.00	8.61	8.98	7.58
Cued	4.39	6.61	[illegible]	6.72	8.33	10.11	7.40	12.44	10.61	11.61	12.39	11.89	10.50	11.57

[a] The abbreviation 1D = one-list delay condition presented to control *Ss*.

nated within a given list condition by providing higher order rhyme cues. In the 6-list condition the facilitating effects of cuing were greater on earlier lists. The *Ss* in the free-recall group recalled an average of 4.89 items from the first list and those in the cued-recall group recalled an average of 12.44 items from this list, a difference of 7.55 items. This difference was gradually reduced, because of decreasing amounts of retroaction in the free-recall groups, until eliminated on the sixth list. The effect was reflected in a reliable Type of Recall × Order of Recall × List Position interaction, F (5, 170) = 3.31, $p < .01$. Fisher's LSD value, based on the pooled between–within error term, was 2.60. It should be noted that although cuing eliminated retroaction within a given list condition, comparison of the column means for cued recall indicated that retroactive effects were still present. Whereas Tulving and Psotka found that cuing restored recall to its original levels, the levels attained in this experiment were still below those of immediate recall.

Since the categories measure showed essentially the same pattern of results as the words measure, a table parallel to Table 2 was not presented. Instead, for all 3 measures, performance was pooled across list position, and the data were analyzed with number of lists and type of recall as between-*Ss* factors and with order of recall as a within-*Ss* factor. The results of these separate analyses indicated that the Type of Recall × Order of Recall interactions were significant for words, F (1, 136) = 95.80, categories, F (1, 136) = 199.58, and words/category, F (1, 136) = 28. 31, $ps < .01$. These interactions are displayed in Table 3. Fisher's LSD values for between-*Ss* comparisons for each of these respective measures were 1.20, .29, and .18; for within-*Ss* comparisons, these values were .46, .18, and .08. The patterns of mean errors for words and for categories indicated that, for both within-*Ss* and between-*Ss* comparisons, significantly more items were recalled on the second recall test when recall was cued. These

TABLE 3

MEAN RECALL OF WORDS, CATEGORIES, WORDS/CATEGORY, AND INTRUSIONS/CATEGORY ON THE FIRST AND SECOND TOTAL RECALL TESTS

Measures	Order of recall tests			
	First		Second	
	Free	Cued	Free	Cued
Words	10.27	9.70	10.53	13.21
Categories	3.89	3.58	3.89	5.41
Words/category	2.53	2.62	2.65	2.42
Intrusions/category	.37	.38	.42	.64

facilitating effects were somewhat greater in the 3- and 6-list conditions than in the ID and 1-list conditions. The *F*s for the 3-way interactions were F (3, 136) = 2.50, $p < .08$, and F (3, 136) = 3.16, $p < .05$, respectively, for the words and categories measures. For the words/category measure, the means presented in Table 3 show that the directions of these effects were reversed. Significantly *fewer* words/category were recalled on the second test when recall was cued. Unlike conceptual category cues, rhyme cues failed to generate the majority of correct instances stored within particular categories. These results suggested that the retroaction displayed in Table 2 was the result of the loss of higher order units as well as the apparent loss of instances within those units. The increment in words recalled under cued recall, and the consequent elimination of retroactive inhibition within a given list condition, apparently occurred because *Ss* were recalling words from more categories, not because the rhyme cue provided access to the majority of correct instances within a given category.

This failure of the cue may have resulted because of difficulty encountered in discriminating intralist and extralist instances within rhyme categories (e.g., Bower, Clark, Lesgold, & Winzenz, 1969). If this supposition is correct, then extralist intrusions per category ought to be greater under conditions of cued as compared with free recall. To check on this possibility extralist intrusions/category were tabu-

lated for recalled categories (i.e., at least one correct instance had to be recalled from the category). Nearly all *S*s exhibited these intrusions. The results of the statistical analysis indicated that the Type of Recall × Order of Recall interaction was significant, F (1, 136) = 9.96, $p < .01$. This interaction effect, which is displayed in the bottom row of Table 3, indicated that significantly *more* intrusions/category were obtained under cued-recall conditions. Sixteen percent of the items recalled per category in the cued-recall conditions were intrusions. Fisher's LSD values for within-*S*s and between-*S*s comparisons were, respectively, .08 and .13. These findings suggested that the rhyme cues provided access to the category as a unit, and that the loss of instances within the unit may have been produced by the difficulty of discriminating between intralist and extralist instances within that unit. In fact, pooling the results for the words/category and intrusions/category measures indicated that items recalled per category was a constant of approximately 75% for all conditions, a value similar to that reported by Tulving and Psotka (1971). An analysis of variance of these pooled measures indicated that none of the sources of variance was statistically reliable, indicating that items/category was relatively invariant.

Discussion

The results of this experiment can be described by summarizing 3 major findings. First, the physical properties of words, in this case their visual-phonetic characteristics, can be used to subjectively organize words already represented in memory. That is, rhymes, like conceptual categories, apparently can function as higher order memory units. Second, the retroactive inhibition in free recall of lists organized into rhyme categories represents a decrement in the recall of these higher order units as well as instances within recalled units. Retroaction effects on the first list were evident within all measures of recall performance: words, categories, and words within recalled categories. Finally, these retroaction effects reflect both inaccessibility of higher order units and unavailability of instances within those units. Providing retrieval cues in the form of rhyme endings provided efficient access to stored information, essentially eliminating retroactive effects. However, retention was not restored to the level attained in immediate recall, and the recall of correct instances within recalled categories actually decreased when relevant cues were provided.

The findings associated with the loss of rhyme categories and the subsequent elimination of retroaction when these higher order units were provided at recall parallel findings associated with conceptual categories as reported by Tulving and Psotka (1971). The losses obtained for instances within recalled categories contrast with the invariance obtained for this measure when instances belong to conceptual categories. In the case of rhymes, instances appear to have been lost from the store since they are "unavailable" even when a highly relevant cue has been provided to aid their retrieval. This loss may have occurred because of the difficulty of making discriminations between prescribed and extralist instances. The rhyme cues are presumed to be effective higher order retrieval aids providing access to the category of relevant words. This effectiveness may (Tulving & Thompson, 1973) or may not (Bahrick, 1970) be determined by what information is encoded during the study trial. However, it is assumed that the cue initiates a search through relevant instances. For each retrieved instance a decision is made as to whether the item was or was not from the list (Anderson & Bower, 1972). The *S*s reportedly experienced little difficulty in generating instances from the category cue and many indicated that they simply matched word beginnings to the rhyme endings, often by going systematically through the alphabet. Accordingly, what is apparently lost from storage is the ability to remember what items of the category appeared in the list and what items did not. This explanation accounts adequately for the facilitating effects of cuing with word endings and for the extralist intrusion findings associated with the experiment. To account for the invariance in words/category obtained with conceptual categories and for its absence with phonetic categories, it must be assumed that greater relative difficulty is encountered in differentiating instances defined by rhyme than in

differentiating instances defined by conceptual relations. A logical analysis of the 2 types of "categories" attests to the reasonableness of this assumption. Each instance of a rhyme category literally recapitulates the category itself, e.g., "OOT" is contained in BOOT, ROOT, etc. Hence, all instances are directly part of the category and, for that matter, all instances are directly related to each other since, by definition, to be a member of the category requires the sharing of identical physical features. Each exemplar of a conceptual category is associatively related to the category name but recapitulates it only by virtue of the strength of this associative relationship and, furthermore, the instances within the category may or may not be associatively related.

One implication of this interpretation is that similar mechanisms may be involved in the acquisition and retention of lists containing conceptually defined categories and lists containing rhyme-defined categories. Whether the results indicate that words/category is invariant may simply be dependent upon the degree of relatedness of the instances falling within the category. If this interpretation is correct, then invariance of words/category may not be obtained if only some portion of the items of an exhaustive conceptual category are used since items within these types of categories tend to be highly related (e.g., Cohen, 1963).

REFERENCES

Anderson, D. R., & Bower, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, **79**, 97–123.

Bahrick, H. P. A two-phase model for prompted recall. *Psychological Review*, 1970, **77**, 215–222.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 481–493.

Cohen, B. H. Recall of categorized word lists. *Journal of Experimental Psychology*, 1963, **66**, 227–234.

Gibson, E. J. Perceptual learning and the theory of word perception. *Cognitive Psychology*, 1971, **2**, 351–368.

Marshall, G. R. Stimulus characteristics contributing to organization in free recall. *Journal of Experimental Psychology*, 1967, **6**, 365–374.

Tulving, E., & Psotka, J. Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology*, 1971, **87**, 1–8.

Tulving, E., & Thompson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352–373.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

Wood, C. *Wood's new world rhyming dictionary.* New York: World Publishing, 1971.

(Received May 10, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 284–290

CODING AND VARIED INPUT VERSUS REPETITION IN HUMAN MEMORY[1]

HENRY C. ELLIS,[2] FREDERICK J. PARENTE, AND CRAIG W. WALKER

University of New Mexico

Two experiments examined the effects of constant vs. varied input of letter strings on recall and recognition memory. Letter strings constructed from pairs of trigrams were fractionated or spatially grouped and presented as either a constant or varied grouping on successive presentations. Varied input produced much greater recall of the entire letter sequence than did constant input. Presumably, *S*s in the varied condition were more likely to adopt a strategy of looking for and encoding a larger unit or structure than *S*s in the constant condition, a strategy which facilitates recall. In contrast, recognition was enhanced by constant input of the letter sequences, a finding predictable from simple frequency considerations.

The following experiments concern the role of nominal-stimulus variation vs. constant or fixed input on recall and recognition memory for groupings of letter sequences. A classical assumption in human memory has been that repetition of stimulus events, in the sense that nominal stimuli are presented in the same manner or form on successive study trials, results in superior memory compared to conditions in which nominal stimuli are varied in some fashion. Nominal-stimulus variation has generally referred to situations in which contextual and/or structural variations of stimuli are introduced in the course of learning.

This assumption has been recently challenged by Bevan and his colleagues in several experiments which have shown that variation among instances of stimulus categories could increase the likelihood of recall. In one study (Dukes & Bevan, 1967), stimuli of the class ax, bx, cx, dx, etc., were presented, where x was the common feature and a, b, c, and d were the modifying elements. For example, x represented a photograph of a particular person, and a, b, c, or d represented his appearance in different poses or dress. To each stimulus compound of a given class *S*s learned a particular name response and then were tested for recall of that response when shown either one of the training stimuli (e.g., bx), or new instances of the category (e.g., jx). They found that the superiority of varied training depended upon whether the test stimulus was a truly old one vs. a "new" item from the category employed. Specifically, when the stimuli were varied during training, varied training produced superior cued recall (using either old or new stimulus items) compared with repetition training (using only new stimulus items). In contrast, repetition training always produced superior recall compared with all other conditions when the test stimuli were old.

Although these results place constraints on any simple generalization about the superiority of varied training, a similar study (Bevan, Dukes, & Avant, 1966) using free-recall procedures indicated that the free recall of generic stimuli (e.g., "pine tree," "oak tree") increased as a result of varied context with both pictorial and word stimuli. Nevertheless, the superiority of varied training was quite small, although reliable, and attempts by Goggin to replicate this finding with word stimuli have not been successful (J. Goggin, personal communication, March 1971).

An alternative approach (e.g., Bower & Winzenz, 1969) is to vary directly the structural properties of the *entire* stimulus to be placed in memory as distinct from varying the modifying elements. The

[1] This research was supported by National Science Foundation Grant GB-27413X to the first author.

[2] Requests for reprints should be sent to Henry C. Ellis, Department of Psychology, University of New Mexico, Albuquerque, New Mexico 87131.

tudies by Bevan (Bevan et al., 1966; Dukes & Bevan, 1967) have varied only modifying features, making them in sense concept identification studies cause a common relevant cue is always ent in the stimulus compound; the ulus context is varied, whereas the be-recalled response term is kept con- nt. This distinction differentiates be- en 2 kinds of variation, which may e different effects on memory: variation posed on the entire event or sequence be remembered vs. variations in only contextual stimuli present during t aining.

The purpose of the present experiment was to exploit the second approach by varying the grouping structure of pairs of trigram doublets (e.g., DENGIT) and to determine the effect of constant vs. varied presentation of grouping structures on recall and recognition. Bower and Winzenz (1969) have reported that variation of the grouping structure of digit series presented auditorily leads to severe degradation of recall and recognition performance. Only when the digit series were grouped in a constant fashion on successive trials did improvement in memory result. The digit series lacked, however, any evident "inherent" overall structure that might be detected by *S*, since they were random permutations of digits. This is simply to say that a string such as *391*, *87*, *625* is no better a 3-chunk set than is *39*, *1876*, *25*, or vice versa; there is no inherently better combination of the numbers for *S* to seek. In this instance, stability rather than variability of the *E*-imposed grouping structure was essential for memory improvement.

It is reasonable to examine the effect of constant vs. varied input under conditions in which an existing structure is present in the series. Structure is used here simply to refer to natural groupings of letters, pronounceable groupings, or easily integrated units. If an inherent (preexperimentally developed) structure is present, as well as detectable by *S*, then it is conceivable that varied input, in the sense of changes in grouping structure on successive presentation, may facilitate recall of the letter series. This finding could result from the modification of *S*'s search strategy in looking for groupable letter sequences. This is simply to suggest that varied input may lead *S* to achieve *S*-imposed encodings, based upon the existing overall structure, more rapidly than would occur with constant input. In contrast, *S*s receiving constant input of letter sequences may be more likely to process them as presented by *E*, and in so doing, they would be less likely to detect codings that would enable a high level of recall. Trigram doublets such as CUPNET or CIMBUX contain an easily recallable structure, if detected, and thus the recall of a series of letters, even when "fractionated" according to different grouping rules on successive presentations, may be enhanced by varied grouping of the sequence. Constant input in this experiment refers to a given doublet such as CUPNET being presented, for example, as C UPN ET consistently on successive presentations, whereas varied presentation refers to changes in the grouping structure on successive presentations, such as C UPNE T on one and CU PN ET on another.

Experiment I examined the effects of constant vs. varied input on recall of letter groupings and indicated that varied input produced superior recall. Experiment II examined the effects of input structure on recognition memory and revealed that constant input produced superior recognition, a finding predictable from simple frequency considerations. In both experiments, 3 levels of trigram meaningfulness (*m*) were employed in order to examine the effects of varied–constant input as a function of *m*.

Experiment I

Method

Design. The *S*s were 30 undergraduate student volunteers from the University of New Mexico. Each *S* received 8 study–test trials to learn and recall a mixed list of 12 trigram doublets. A constant and a varied condition of stimulus presentation were employed. In the constant condition, the doublets were presented consistently on each study trial. In the varied condition, *S*s saw different perceptual groupings of each doublet on successive lists presented within a trial. Inter-

spersed between each study trial was a test trial in which S was required to recall the entire sequence of 6 letters in the order and sequence in which they were presented.

Stimuli. The stimulus items were trigram doublets constructed by combining 2 trigrams of comparable m values selected from the Archer (1960) norms. For example, the trigrams RUM and WIG were combined to form the doublet RUMWIG. The average association values for the 4 high-, moderate-, and low-m trigram doublets were 100%, 56%, and 12%, respectively. Doublets used were RUMWIG, DENGIT, BANCOW, CUPNET, CIMBUX, JUMHYR, PUZWEK, SUBTYC, TIJFEQ, NEJWOQ, YIWFEP, and MOJZUV. Each of the doublets was fractionated (spatially grouped) into 3 groupings of letters according to one of 4 grouping rules. For example, the 4 high-m doublets, RUMWIG, DENGIT, BANCOW, and CUPNET, once fractionated were presented as groupings of R UMWI G, DE NGI T, BA NC OW, and C UPN ET, respectively. The 4 moderate-m and 4 low-m stimuli were also fractionated according to the same rules.

Four stimulus lists were prepared, each list containing the 12 fractionated stimuli presented in a different serial order. In the constant condition, each doublet was presented in the same fractionated fashion over successive presentations of the lists. In the varied condition, the doublets were fractionated according to a different rule over successive presentations of the 4 lists. For example, the doublet BANCOW appeared as B ANCO W, B ANC OW, BA NC OW, and BA NCO W, on each list, respectively. Thus, the grouping rules specify letter groupings as follows: 1, 4, 1; 1, 3, 2; 2, 2, 2; and 2, 3, 1. Variation of the nominal stimulus was thus defined in terms of the use of a different fractionation rule applied to the same doublet on successive list exposures. Moreover, the intact trigram (e.g., BAN or COW) was never presented in either the varied or constant condition. The letters of a given doublet were always presented in the same order and sequence for both conditions, despite changes in the grouping structure for the varied condition.

Procedure. There were 8 study–test trials administered in each condition, a trial consisting of the successive presentation of 4 stimulus lists of the 12-doublet letter strings. Thus, S saw a total of 48 (4 Lists × 12 Doublets) stimulus presentations per trial. Each fractionated doublet was exposed for a 2-sec. interval on a Stowe memory drum, and S was instructed to remember the items, as he would be tested for recall. Since S saw each doublet fractionated 4 ways in the varied condition, input variation occurred *within* each trial; on successive trials, the same 4 varied versions were repeated.

The Ss were given a recall task interspersed between each trial that required them to recall the letters in the order and sequence presented. The Ss were given a test sheet containing 12 horizontal arrangements of 6 adjacent boxes, and S filled in each arrangement of boxes with the appropriate stimulus letters. The Ss were i[illegible] to fill in all of the boxes (guessing unknown [illegible] whether or not they were sure of their r[illegible] making the procedure one of forced reca[illegible] though the task required that the letters b[illegible] in the order and sequence in which th[illegible] shown, Ss were allowed to free recall the let[illegible]

Results

Recall. The principal data obtai[illegible] in recall are shown in Figure 1, whic[illegible] a plot of the probability of correct com[illegible]ete recalls (all 6 letters in correct ord[illegible]nd sequence) for the constant and vari[illegible] input conditions over 8 trials. The [illegible]ure shows a substantial superiority in [illegible]ecall due to varied input throughout the test sequence, F (1, 28) = 12.90, $p < .005$. Terminal recall probabilities for the v[illegible]ried and constant conditions were .73 and .43, respectively. Recall improved for [illegible]oth conditions as a result of practice, F (1, 196) = 81.08, $p < .001$, and the rate of improvement was faster under the va[illegible]ied condition, as reflected in the significant Type of Training × Trials interaction, F (7, 196) = 10.89, $p < .001$. The effect of meaningfulness was also significant, F (2, 56) = 58.73, $p < .001$. Mean probabilities of recall in the varied condition for the 3 levels of meaningfulness (high, moderate, and low) were .70, .45, and .4[illegible], respectively, whereas the comparable means of the constant condition were .41, .21, and .11, respectively. Since meaningfulness did not interact with type of training, F (2, 56) = .67, $p > .20$, separate recall figures showing this effect are not shown. The data of Figure 1 are thus based upon all levels of meaningfulness.

Partial recall of sequences. Partial recalls of 5-, 4-, 3-, and 2-letter strings were calculated in order to examine the effect of varied-constant input on size of the recalled chunk. A partial recall was defined as the correct recall of 2 or more *adjacent* letters. Thus, a correct partial recall of a 3-letter sequence (unit) for the trigram RUMWIG would be, for example, RUM, UMW, MWI, or WIG. A given recalled sequence was scored only once, so there were no duplicate or nonindependent scorings. For example, a partial recall of RUMW was

·d as 1 4-unit recall, not as 2 3-unit lls or 3 2-unit recalls. As expected, varied group gave more 4- and 5-unit ings, F (1, 14) = 11.82, $p < .005$, reas the constant group gave more and 3-unit strings, F (1, 14) = 52.2, .001. This trade-off is necessary, ce the higher recall of 2- and 3-unit ings for the constant group is due to fact that the varied-group *S*s use up eir potential 2- and 3-unit strings in eir more frequently recalled 4- and 5-unit ·ings.

Single-letter recall. A third recall meaure, the total number of correct singletter recalls, was calculated. Single-letter ecalls were again scored only if they were n the correct serial position, and each recall was scored correct or incorrect regardless of the recall of an adjacent letter. Thus, on a given trial there were 72 possible correct single-letter recalls (12 Trigram Doublets × 6 Letters per String). The mean correct single-letter recalls, averaged over 8 trials, were 49.9 for the varied condition and 31.3 for the constant condition, F (1, 238) = 62.44, $p < .001$. This simply indicates that the superiority of the varied condition in recall holds regardless of whether single letters or the entire sequence of 6 letters is the scoring unit.

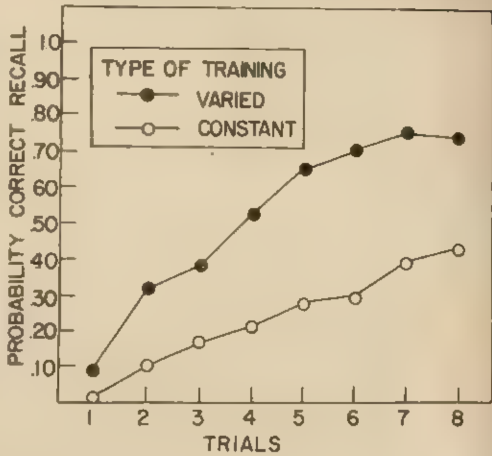


FIGURE 1. Probability of complete recall of the letter sequences as a function of trials, with varied vs. constant input of letter groupings as parameter.

EXPERIMENT II

The second experiment examined the effects of varied vs. constant training on recognition memory where it was proposed that nominal-stimulus variation would lead to poorer recognition performance.

Method

Design. The *S*s were 34 undergraduate student volunteers from the University of New Mexico. The *S*s were given training with varied or constant presentations as in Experiment I, followed by a 2-item forced-choice recognition test. Three modifications of the training procedure were introduced. The study trial consisted of only one presentation of the list (rather than 4), followed by the recognition test. Thus, in the varied condition, *S* saw a different list on each trial for a block of 4 trials. Twelve study trials were administered (rather than 8), with lists varied within blocks of 4 trials. Thus, *S*s in the recognition experiment were given 12 presentations/trial for a total of 144 presentations (12 Items × 12 Trials), whereas *S*s in the recall experiment were given 48 presentations/trial for a total of 384 presentations (4 Lists × 12 Items × 8 Trials). This was done simply because it was unnecessary to provide as much training in the recognition experiment. Each fractionated doublet was presented for only 1 sec.

Procedure. The experiment again used a study–test procedure. After each study trial, *S*s were shown the old stimuli singly, paired with a highly similar distractor item. Distractors were formally similar to the old (target) items, with a distractor consisting of a 1-letter substitution of one of the 4 middle letters of the old doublet. A vowel was replaced by a vowel and a consonant was replaced by a consonant. For example, the old item R UMWI G was paired with the distractor R AMWI G on one trial and R UMNI G on another trial, for the constant condition. Thus, an old item and its distractors retained the same fractionation rule on study and test trials in the constant condition. A different distractor item was used on each of the 12 trials, and the right–left position of the pairs was randomly varied on each trial. The *S*s were instructed to underline whichever item in the pair they had seen during the study trial. The procedure was the same for the varied condition except that the distractor items were fractionated according to the same rule as the old item as it appeared on a given study trial. Thus, on a given recognition test, R AMWI G was paired with R UMWI G, and on a subsequent test, RU MN IG was paired with RU MWI G.

Results

Recognition. The principal data obtained in the recognition test are shown

in Figure 2, which is a plot of probability of correct recognition over test trials. This figure shows that constant training leads to superior recognition compared with varied training, $F\ (1,\ 32) = 15.16$, $p < .001$. Terminal probabilities of correct recognition for the constant and varied conditions were .90 and .80, respectively. Recognition improved for both conditions as a result of training, $F\ (11, 352) = 14.54$, $p < .001$; however, there was no reliable Type of Training × Trials interaction, $F\ (11,\ 64) = 1.20$, $p > .20$. The effect of meaningfulness was also significant, $F\ (2,\ 252) = 38.18$, $p < .001$. Mean recognition performances in the constant condition for the 3 levels of meaningfulness (high, moderate, and low) were .85, .85, and .73, respectively, whereas the comparable means for the varied conditions were .75, .73, and .63 respectively. Again, since meaningfulness did not interact with type of training, $F\ (2, 64) = 1.23$, $p > .20$, only a single recognition plot is presented.

Discussion

The principal findings of the experiments were that varied input, in the form of altered groupings of letter sequences taken from trigram doublets, facilitated recall of the letters, whereas constant input facilitated stimulus recognition. The superiority of varied input on recall occurred whether measured by complete recall of all letters in the appropriate position or by total number of correct single-letter recalls. Thus these data indicate one restriction on any assumption that sheer repetition of stimulus events is the condition for achieving maximum recall.

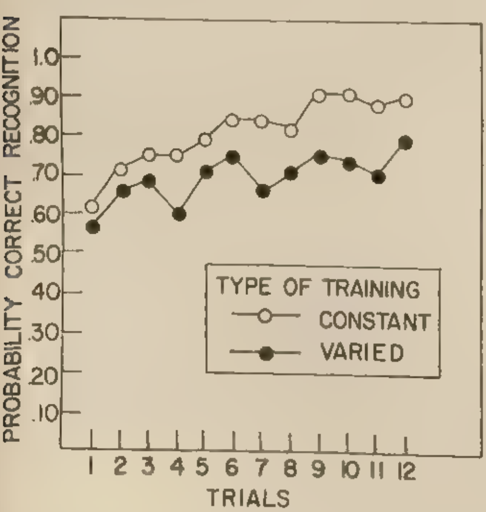


FIGURE 2. Probability of correct recognition of letter sequences as a function of trials, with varied vs. constant input as parameter.

An explanation of the general superiority of varied input on recall is reasonably straightforward and can be accounted for by what we call a *perceptual regrouping hypothesis.* The process is viewed as operating in the following manner. When S is shown varied groupings on successive presentations, he eventually learns to ignore the E-imposed groupings, since they change on successive presentations and are not a relevant feature. In addition, it is assumed that varied input leads S to adopt the strategy of seeking a new and overall grouping which, if successful, will facilitate recall because the particular letter sequences used can be grouped as a unit. Since the letter sequences contained an inherent grouping structure in the form of pronounceable or easy-to-integrate units, detection and encoding of this overall structure must facilitate recall. Moreover, the recall data suggest the general principle that variability in nominal-stimulus input leads to greater stability in S's functional encoding under conditions in which S can achieve an overall or unitary encoding of the input sequence. In contrast, if no existing or inherent structure were present, varied input would not be expected to facilitate recall. Search for such a structure and the corresponding development of S encodings would be fruitless, and performance in such instances would be degraded.

On the surface, the superiority of recall under varied input appears contrary to the results reported by Bower and Winzenz (1969) and may further imply that their reallocation hypothesis is limited. Critical differences between the experiments appear, however, to make the superficially divergent results understandable. The perceptual regrouping hypothesis can account for differences between this experiment and Bower and Winzenz's Experiments I and II regarding recall, where in the latter study it was necessary for S to have a stable E-imposed structure for learning to occur. Moreover, a stable E-imposed structure was necessary for learning even when letter sequences were used (Winzenz, 1972, Experiment V), if the groupings themselves

ıf words. The perceptual regrouping ьesis assumes only that varied input in m of letter groupings, instead of being .ed to separate storage locations, even- ·omes to be regrouped as an integrated nd is then shunted to the same storage on. The point is that *S* must discover a encode a sequence that contains an in- h .nt structure if variable input is to enhance

ere are other differences between these riments, including length of series and ınt of information processed prior to each In addition, differences in auditory vs. ıal presentation may have contributed to pattern of differences in recall between 2 studies. In the present experiment, saw the entire letter sequence and would esumably have greater opportunity to en- ode it as an integrated unit. In contrast, wer and Winzenz (1969) presented the git series auditorily, either as single digits ing read with pauses between groups, or as successive numerical groupings being named. n either instance, auditory presentation ould appear to reduce the likelihood of *S* detecting a larger structure or grouping even if such structure had been present. In short, although there is no doubt that the consistency of *E*-imposed groupings is necessary for improvement in recall under conditions examined by Bower and Winzenz, structural variability of letter sequences aids recall under conditions in which *S* can detect and encode an underlying structure consisting of the entire unit. More generally, these findings further attest to the significance of encoding processes in human learning and memory and to the importance of structural properties of stimuli in learning tasks (cf. Ellis, 1973).

The superiority of recognition under constant input is predictable from relatively simple frequency considerations. The recognition task does not require integration or organization of the letter string into a larger unit. Moreover, the test compared distractors conforming to the same grouping rule as the old items, so that *S* did not have to attend to or encode the particular grouping in the constant condition. Thus, the variable condition reduces the likelihood that *S* will develop a stable encoding *necessary for recognition.* Since the recognition task requires a choice between 2 grouped sequences, *S* would be less disposed to develop unitary encodings in the recognition task. All that would appear to be necessary for correct recognition is the development of a relatively stable encoding which is maximized under conditions of consistent presentation of a fixed structure.

The positive relationships between meaningfulness and recall and between meaningfulness and recognition (Ellis, Parente, & Shumate, in press; Ellis & Shumate, 1973) is well-established and is of no great interest here. What is of interest is the lack of an interaction between varied–constant input and meaningfulness in recall. Although the superiority of the varied condition occurred across all levels of meaningfulness, it might have been expected that the effect would have been attenuated at lower *m* levels, since these items are less easily unitized. Since this did not occur, it would seem reasonable to assume that even low-*m* trigram doublets possess sufficient inherent structure for the variability effect to appear.

Finally, these experiments indicate that variably grouped presentations of letter sequences affect recall and recognition differentially. This outcome is not readily handled by a strength theory of recall and recognition (cf. Kintsch, 1970), which assumes that recall and recognition involve essentially the same processes. The fact that varied input affects recall and recognition in different ways suggests that different processes may underlie the 2 performances. Some caution must be used in making this inference, however, because the 2 training procedures were not exactly comparable, with different amounts of practice, different stimulus durations, and differences in manipulating variability (within trial in recall and between trials in recognition). Despite these differences, the effect of variability was reversed for the 2 conditions over the entire training testing sequence.

REFERENCES

ARCHER, E. J. A re-evaluation of all possible CVC trigrams. *Psychological Monographs*, 1960, **74**(10, Whole No. 497).

BEVAN, W., DUKES, W. F., & AVANT, L. The effect of variation in specific stimuli on memory for their superordinates. *American Journal of Psychology*, 1966, **79**, 250–257

BOWER, G. H., & WINZENZ, D. Group structure, coding, and memory for digit series. *Journal of Experimental Psychology*, 1969, **80**(2, Part 2).

DUKES, W. F., & BEVAN, W. Stimulus variation and repetition in the acquisition of naming responses. *Journal of Experimental Psychology*, 1967, **74**, 178–181.

Ellis, H. C. Stimulus encoding processes in human learning and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*. Vol. 7. New York: Academic Press, 1973.

Ellis, H. C., Parente, F. J., & Shumate, E. C. Meaningfulness, perceptual grouping, and organization in recognition memory. *Journal of Experimental Psychology*, in press.

Ellis, H. C., & Shumate, E. C. Encoding effects of response belongingness and stimulus meaningfulness on recognition memory of trigran muli. *Journal of Experimental Psychology*, , 98, 70–78.

Kintsch, W. Models for free recall an ogni- tion. In D. A. Norman (Ed.), *Models uman memory*. New York: Academic Press,

Winzenz, D. Group structure and codi serial learning. *Journal of Experimental logy* 1972, 92, 8–19.

(Received May 15, 1973)

[illegible] [illegible]erimental Psychology
[illegible] No. 2, 291–297

ENCODING AND RETRIEVAL PROCESSES IN LONG-TERM RETENTION

AXEL GÖTZ AND LARRY L. JACOBY [1]

Iowa State University

Retention effects of retrieval practice and of predictability of recall delay were investigated in 2 experiments. Several 5-word lists were recalled initially after no delay or delay filled with number subtraction; delay type was predictable for *S*s in a precue but not in a postcue condition. Words from all lists were tested in final free recall (FFR). The advantage of filled over no delay in FFR was greater in the precue than in the postcue condition. This result was taken as evidence that attributes selected for encoding depended on the type of delay anticipated. In Experiment II only half of the lists were recalled initially. Initial recall aided FFR more for filled- than for no-delay words. It was concluded that retention benefits from initial retrieval to the extent that retrieval cues used at initial and final recall are similar.

[illegible]earsal has been used extensively as a[illegible]lanatory concept in recent analyses o[illegible]ory processes. In Atkinson and [illegible]'s (1968) 2-store theory of memory, re[illegible]l serves the dual purpose of main-ta[illegible] items in short-term store (STS) and o[illegible]ermining the amount of information ab[illegible] an item transferred to long-term st[illegible] LTS). Serial position effects in-clu[illegible]ing the negative recency effect found in final free recall (FFR), have been interpreted in terms of the positive relationship between rehearsal frequency and transfer to LTS (Atkinson & Shiffrin, 1971; Craik, 1970). More recent results suggest, however, that rehearsal might only serve the function of maintaining items in STS (Jacoby, 1973; Jacoby & Bartz, 1972; Meunier, Ritz, & Meunier, 1972). In the Jacoby and Bartz study, 3 groups of *S*s were presented with several 5-word lists; each list was recalled initially, either immediately after its presentation, after a silent-delay interval, or after a delay filled with number subtraction. Initial recall int he subtraction condition was well below that in the other 2 conditions, supporting the contention that rehearsal is necessary to maintain items in STS. However, in an unexpected FFR test for all presented words, recall after silent delays was lower than after number subtraction; amounting to an inverse relationship between rehearsal opportunity and storage in LTS. This and the finding that silent-delay items were not better recalled than no-delay items led to the conclusion that rehearsal frequency does not determine transfer to LTS. As possible explanations for the final-recall advantage of filled-delay items, the authors suggested differential encoding during study and differential retrieval practice at initial recall.

According to the differential encoding interpretation, the coding of items during study was dependent on the nature of delay anticipated. For immediate recall or recall after a silent delay, it would be sufficient to keep list items in STS by means of rehearsal with no necessity to generate more permanent retrieval cues. The *S*s in the subtraction condition were likely to organize items or process them to a "deeper level" (Craik & Lockhart, 1972) to insure that they would be retrievable after the filled delay. Thus, subtraction *S*s generated longer lasting retrieval cues that more likely would be still accessible at the time of FFR. The second interpretation appeals to the influence of retrieval practice on later free recall (e.g., Lachman & Mistler, 1970). Immediate recall or recall after a silent delay might only make use of short-lived retrieval cues such as the acoustic trace

[1] Requests for reprints should be sent to Larry L. Jacoby, Department of Psychology, Iowa State University, Ames, Iowa 50010.

of an item (Tulving, 1968), corresponding to retrieval from STS. In contrast, retrieval after a filled delay might necessitate the use of more permanent retrieval cues. Filled-delay items would thus benefit more from retrieval practice since the retrieval cues employed would be more similar to those accessible at the time of final recall. Since differential encoding is assumed to depend on *S*'s ability to anticipate the nature of the delay interval, removal of this ability should nullify the difference between filled-delay and no-delay items in final recall. Differential retrieval practice would be expected to operate regardless of *S*'s ability to anticipate the nature of the delay preceding initial recall. Predictability of delay preceding initial recall was manipulated in the present investigation in order to separate the effects of differential encoding and differential retrieval.

In Experiment I, short lists were presented and free recalled either immediately or after a period of number subtraction. In a precue condition, *S*s were informed prior to list presentation concerning which of the 2 delay conditions would follow. Thus, they had the opportunity to encode words more permanently preceding a filled-delay interval. Final recall for the precue condition was expected to replicate the earlier reported superiority of filled over no delay. In a postcue condition, information about delay type was not available until after list presentation. Since delay type was not predictable during study, the final recall advantage of filled-delay items should be absent if differential encoding was the only factor responsible for the earlier results. No effect of the cuing manipulation would be expected if only differential retrieval was operative in the earlier experiment.

Experiment I

Method

Materials and procedure. Two sets of 150 words with A and AA ratings were randomly selected from the Thorndike and Lorge (1944) wordbook, and used to form 2 replications of the basic design. Within each set, words were assembled into 30 5-word lists and tape recorded for auditory presentation at a 2-sec. rate. The *S*s' oral recall was recorded by *E*. Half of the lists were recalled immediately (no delay, ND) after their presentation while the other half were recalled after a filled-delay (FD) interval. The sequence of immediate and delayed recall was randomized separately for each of 2 sets of lists. The task employed in FD intervals consisted of auditory presentation of randomly selected 2-digit numbers. Numbers were presented at a 2-sec. rate, with the first number occurring 1 sec. after the last word in a list; 7 numbers were presented within the 15-sec. delay interval. The *S*s were instructed to subtract 1 from each number and say the result aloud prior to the presentation of the next number. For all conditions, the word "go" signaled the beginning of the recall interval; recall was spoken. The 7.5-sec. recall interval was terminated by the word "ready" which preceded the first word of the next list by 2 sec.

The *S*s in a precue condition were informed prior to the presentation of each list concerning the nature of the delay that would follow. A stack of 3 × 5 in. note cards was placed before *S*s in that condition; each list was represented by a card. Printed on each card was either the word delayed or immediate, indicating the nature of delay. Upon hearing "ready" preceding list presentation, *S* read the top card and then placed it next to the stack; the card read corresponded to the list that was to be presented. In a postcue condition, *S*s learned about delay after presentation of each list by hearing either the signal for recall or the first of the 2-digit numbers for subtraction

Following recall of the last list, all *S*s were read 3 sets of 9 digits and asked to recall each set in order. Next, *S*s were instructed to write down all the words they could remember from all lists. Prior to these instructions there was no reason for *S*s to anticipate the FFR test. There was no time limit on the final test. The digit span task was employed as an attempt to minimize the effects of retrieval from STS in FFR.

Subjects and analyses. The *S*s were 20 volunteers from undergraduate courses at Iowa State University; 10 *S*s were assigned to each of the cue conditions according to a prearranged random schedule with the restriction that the ratio of males to females must be constant across conditions. The *S*s were tested individually and received extra course credit for their participation. For each *S*, recalled words were classified according to delay condition and serial position.

Analyses to be reported employed a 2 × 2 × 5 (Cue × Delay × Serial Position) analysis of variance with repeated measures on the last 2 factors; initial and final recall were analyzed separately. Replications were not included as a factor since a preliminary analysis indicated that results did not differ across replications.

Results and Discussion

Initial recall. Recall probability as a function of input position, delay, and cue

condition is displayed in Figure 1. The FD items were recalled at a lower level than ND items, F (1, 18) = 313.67, $p < .001$, indicating that interpolated subtraction interfered with the maintenance of list items in STS. Recall probability was slightly higher in the precue (.74) than in the postcue (.67) condition, F (1, 18) = 4.8[illegible], $p < .05$. The main effect of serial position, F (4, 72) = 14.72, and the Delay × Serial Position interaction, F (4, 72) = 9.[illegible], were also significant ($ps < .001$). Recall probability declined steadily across serial positions in the FD condition while remaining relatively stable in the ND condition.

Final recall. Final recall probability was higher for FD than for ND items, F (1, 18) = 33.[illegible], $p < .001$. As predicted by the differential encoding hypothesis, the recall advantage of FD over ND items was greater in the precue (.18, .06) than in the postcue (.15, .12) condition, F (1, 18) = 9.66, $p < .01$. Overall, recall probability declined across input serial positions, F (4, 72) = 6.51, $p < .001$. Although the appropriate interaction was not significant, the decline across positions was almost totally absent for recall of ND items in the precue condition.

Results of the present investigation provided support for the differential encoding hypothesis. When *S*s could anticipate delay type, FD items were coded in a more permanent fashion than were ND items. Coding of the 2 types of items was necessarily the same when *S*s could not predict delay type; coding when delay was unpredictable appears to have been intermediate to the 2 forms of coding employed in the precue condition. The decline in recall probability across input positions can be attributed to the study of earlier items during the presentation of later ones (e.g., Atkinson & Shiffrin, 1971). The absence of serial position effects when *S*s anticipated an immediate recall test suggests that the cumulative study strategy was primarily used to prepare for delayed recall.

The failure to obtain an effect of delay in the postcue condition can be taken as evidence against the differential retrieval

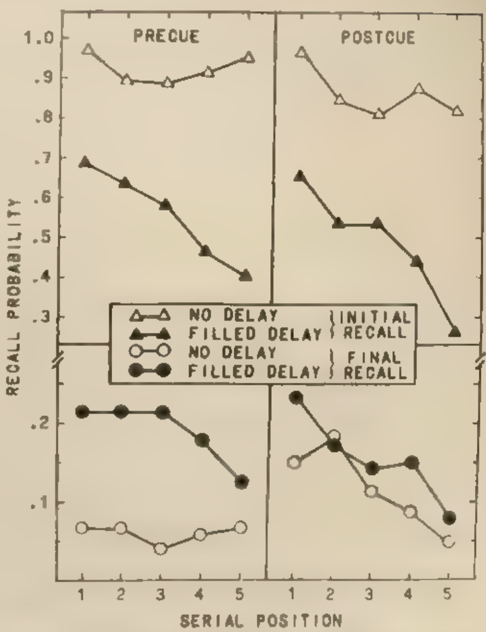


FIGURE 1. Probability of initial and final free recall as a function of serial position, cue, and delay condition in Experiment I.

hypothesis. However, the use of different retrieval cues after the 2 types of delay might develop gradually across the test session. In addition, retrieval cues employed after an FD would at best be expected to be only relatively permanent and might not survive the presentation of a large number of later lists. Both considerations suggest that an advantage for FD items might be more likely in final recall of lists presented near the end of the experimental session. Filled delay and ND in the postcue condition were therefore compared separately for the first 8 and second 7 lists of each type. Final recall probabilities were nearly identical for FD and ND items presented in the first lists studied (.10, .11). However, recall probability from the last study lists was higher for FD than for ND items (.21, .12), t (18) = 2.83, $p < .05$. Thus, the differential retrieval hypothesis cannot be totally dismissed.

Similar comparisons in the precue group reveal higher recall of FD over ND in the first lists studied (.15, .05), t (18) = 3.54, $p < .01$, as well as in the last study lists

(.22, .07), t (18) = 4.78, $p < .001$. Apparently, in contrast to differential retrieval strategies, differential encoding is established early in the course of the experiment.

Experiment II

Experiment II was designed to further separate the effects of differential encoding and differential retrieval. As in Experiment I, *S*s in a precue and postcue condition were exposed to lists followed by either an FD or ND. Now, however, *S*s recalled only half of the lists initially; recall vs. no recall was factorially combined with cue and delay conditions. In all conditions, *S*s were unable to predict whether or not recall would be required. Thus, study encoding should be constant across recall conditions. The effects of differential encoding can be observed uncontaminated by retrieval practice in the no-recall conditions. Comparisons of the final-recall effectiveness of initial recall in the 2 delay conditions yield information that is relevant to the differential retrieval hypothesis. Initial recall after an FD should have a larger effect in FFR if the differential retrieval hypothesis is accurate.

Method

Materials and procedure. A set of 200 words with A and AA ratings were randomly selected from the Thorndike–Lorge (1944) wordbook. These words were assembled into 40 5-word lists and tape recorded for auditory presentation at a 2-sec. rate. The *S*s' oral recall was recorded by *E*. Half of the lists were followed by FD while the other half were followed by ND. Each half was further subdivided into an equal number of lists initially recalled (IR) vs. not initially recalled (NIR). Four delay–recall combinations resulted: FD–IR, FD–NIR, ND–IR, and ND–NIR. Ten lists were randomly assigned to each delay–recall combination; the presentation order of lists was also randomized. Four replications of the basic design were constructed so that each list represented each delay–recall combination equally often. The distractor task employed in FD intervals was the same as in Experiment I, as were the arrangements for recall in the FD–IR and ND–NIR conditions. In the FD–NIR condition, the word "ready" occurred 2 sec. after the last 2-digit number and signaled the beginning of the next list. The word "ready" served the same purpose in the ND–NIR condition but occurred 1 sec. after the last word of the preceding list.

Cuing arrangements for a precue and postcue condition were the same as those in Experiment I. None of the *S*s was precued with regard to recall condition. The recall–no-recall manipulation was explained to *S*s by describing the investigation as being concerned with the effects of disruption on memory. Other procedural details including arrangements for digit span tests and final recall were identical to those in Experiment I.

Subjects and analyses. The *S*s were 40 volunteers enrolled in psychology courses at Iowa State University and received extra course credit for their participation. Twenty *S*s were assigned to each cue condition according to a prearranged random schedule with the restriction that the ratio of males to females must remain constant across conditions.

Study lists were divided into 2 blocks of equal size for preliminary analyses of initial and final recall. The 20 lists of Block 1 were the first 5 lists representing each of the 4 delay–recall combinations while those in Block 2 were the second 5 lists from each combination of conditions. The analysis of initial recall failed to reveal any significant effects of input block. In final recall, the probability of recall was higher for lists presented in the second (.11) than in the first (.06) block, F (1, 266) = 67.94, $p < .001$; effects of other variables were generally consistent across blocks but more pronounced in Block 2. The finding of more pronounced effects in final recall of later presented lists is similar to that reported in Experiment I and may be due to either recency of presentation or the development of study strategies. Since effects in recall of early lists were less clear and did not contribute additional information, results will be reported only for lists presented in Block 2. Initial recall data were treated as in Experiment I. Final recall scores were entered into a 2 × 2 × 2 × 5 (Cue × Delay × Initial Recall × Serial Position) analysis of variance with repeated measures on the last 3 factors.

Since in earlier experiments, including Experiment I, interactions of *S*s with other factors did not reach significance ($p = .05$), it was decided a priori to use the pooled interactions with *S*s as the denominator for F tests. The appropriateness of this procedure was checked for the analyses of initial and final recall. Only 1 of the 8 interactions involving *S*s exceeded the .10 level, F (38, 152) = 1.39, indicating that the assumptions necessary for pooling were met.

Results and Discussion

Initial recall. Initial- and final-recall probabilities for each combination of conditions are plotted in Figure 2. As expected, recall was lower in the FD than in the ND condition, F (1, 342) = 336.70, $p < .001$. The Delay × Serial Position interaction was also significant, F (4, 342)

= 3.7[illegible], $p < .05$. Recall probability declined across input positions in the FD condition while showing a slight increase for ND items. Intrusion errors were infrequent; intrusions from lists that were not recalled were only slightly more frequent ([illegible]2.2%) than those from recalled lists ([illegible].7%). Thus, it appears that *S*s successfully differentiated lists.

Final recall. The effects found in Experiment I were replicated. As shown in Figure [illegible] recall probability declined across serial positions, F (4, 722) = 15.20, p < [illegible] and was higher for FD than for ND [illegible] F (1, 722) = 18.59, $p < .001$. The recall advantage of FD over ND items was larger in the precue (.14, .08) than in the postcue (.12, .11) condition, F (1, 7[illegible]) = 9.68, $p < .01$, again supporting the differential encoding hypothesis. For lists not initially recalled, final recall in the precue condition was higher for FD than for ND items (.10, .05), t (722) = 2.70, $p < .01$. Thus, the effect of differential encoding manifested itself in pure form. Lists that were recalled initially showed higher final recall than did lists that were not initially recalled, F (1, 722) = 29.20, $p < .001$. In addition, the differential retrieval hypothesis predicts that FD items should benefit more from initial recall than should ND items. The corresponding Initial Recall × Delay interaction was significant, F (1, 722) = 4.53, $p < .05$, and showed that initial recall was most beneficial for FD items. As shown in Figure 2, the advantage of FDs was larger in the precue than postcue condition but present in both among items that had been recalled initially. The present investigation thus provided evidence of both differential encoding and differential retrieval.

The results are also relevant to a third possible interpretation of the final-recall advantage of FD items. It might be argued that the subtraction task was not demanding enough to completely eliminate rehearsal, and that the higher final recall of FD items was produced by this additional rehearsal. This interpretation would predict higher recall of FD items even for lists that were postcued and not recalled initially; neither differential encoding nor differential retrieval would predict an effect of delay for these lists. The results showed that final-recall performance was nearly identical for FD and ND items (.09, .10) in the postcue condition that had not been tested initially. Thus, there was no evidence that rehearsal during subtraction aided final recall.

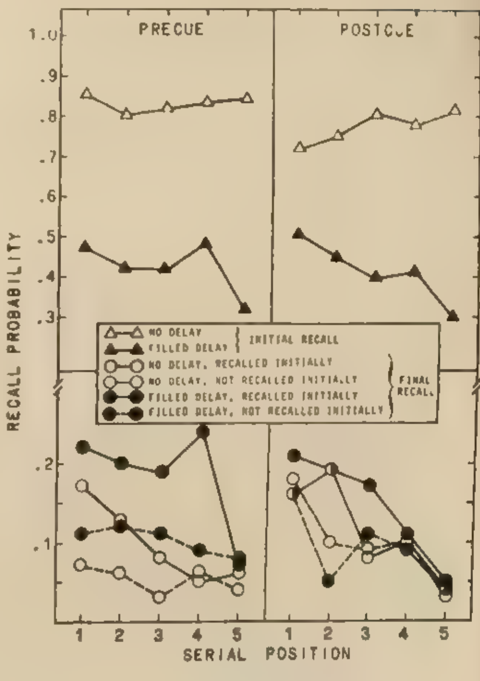


FIGURE 2. Probability of initial and final free recall as a function of serial position, cue, delay and initial recall condition in Experiment II.

General Discussion

Among the most widely observed effects in studies of memory is the dependence of recall probability on study opportunity (e.g., Cooper & Pantle, 1967). One possible explanation of this effect is that rehearsal frequency determines the probability of delayed recall or, in the context of a 2-store theory, the transfer of information to LTS. While the term "rehearsal" has been defined in a variety of ways, it is used in this discussion to denote *S*s' overt or covert repetition of items.

Memory theorists holding widely divergent orientations (e.g., Atkinson & Shiffrin, 1968; Underwood, 1972) have agreed concerning

the central role played by rehearsal frequency. The rehearsal frequency interpretation also has the advantage of parsimony and has been applied to a wide variety of phenomena. Despite its wide acceptance, there is no firm evidence of a causative relationship of rehearsal frequency to delayed-recall probability. The most convincing evidence has come from experiments that have required overt rehearsal and have related rehearsal frequency to recall probability (Atkinson & Shiffrin, 1971; Rundus, 1971; Rundus & Atkinson, 1970). It should be recognized, however, that overt rehearsal only renders observable *S*s' vocalization of study items; rehearsal might be accompanied by considerably more complex processing that is not observable. The relationship of overt rehearsal frequency and recall probability that has been found might be due to the correlation of both factors with underlying coding processes.

Recent evidence suggests that explanations of long-term memory effects in terms of rehearsal frequency are insufficient. Several studies (Jacoby, 1973; Jacoby & Bartz, 1972; Meunier et al., 1972) have found no effect of variations in rehearsal opportunity or frequency of overt rehearsal in long-term memory performance. Moreover, the present investigation found differences in final recall for conditions that were equated with regard to rehearsal opportunity during study; final recall in the precue condition was higher for lists followed by FD rather than ND. It might be argued that the subtraction task employed did not eliminate rehearsal so that the FD condition had additional opportunity for rehearsal. However, further analyses revealed that rehearsal was either absent during delays or did not aid final recall. These results can be accounted for if rehearsal, defined as repetition of items, is assumed to be only one of several types of processing available to *S*s. Rehearsal may be sufficient to maintain items for immediate recall while additional processing is required to aid performance on a long-term memory test (Craik & Lockhart, 1972; Jacoby, 1973).

The trace resulting from the encoding of an item can be conceptualized as a set of attributes, with the level of processing determining the particular attributes included in the set (Craik & Lockhart, 1972). Processing in addition to rehearsal will lead to the inclusion of attributes that are more resistant to loss. If an item is studied under the expectations of immediate recall, its memory trace will contain fewer and less permanent attributes than if it is studied with the anticipation of delayed recall. More permanent attributes are encoded, including organization of list items, when delayed recall is anticipated. Consistent with the preceding statement are *S*s' descriptions of their study activities. When asked after the experiment, several *S*s in the precue condition reported making up sentences, stories, or mental images involving list items when an FD was anticipated, and merely repeating the items when they expected an immediate-recall test. The nature of delay preceding recall was not predictable in the postcue condition. The *S*s in that condition may have either adopted an intermediate level of processing for all lists or vacillated between rehearsing items and attempts to organize them. It should be noted that the relating of processing to *S*s' expectations and cognitive states is not a new development, and gains support from an older literature. Müller (1911, pp. 11–20) listed a number of variables that he found to influence study strategy; the time interval between study and test was among those listed.

Data from the present experiments also provided support for the differential retrieval hypothesis. The final-recall effects of initial recall can be interpreted within the memory attribute framework. Attributes used as retrieval cues experience an increase in cue effectiveness as a consequence. The use of short-lived attributes requires less effort so that they are employed as retrieval cues when there is a choice; immediate recall primarily uses short-lived attributes such as the acoustic trace of an item. If an FD precedes recall, short-lived cues will not be available and initial recall will be based on more durable attributes. Later recall will benefit from initial recall only to the extent that both can use the same attributes as cues. Initial recall after an FD was more advantageous, since the attributes employed for retrieval were more likely to be also available at the time of final recall, than were those used in the ND condition. It would be predicted that the effectiveness of initial recall could be even further enhanced by increasing the duration of the FD interval.

Recall differences observed in the present experiments cannot be explained in terms of amount of study opportunity or rehearsal frequency. Rather, recall performance appears to be a function of the attributes that are accessible at the time of test. Accessi-

bilit[illegible] depends on 2 factors: (*a*) Dif-fere[illegible]oding during study determines whi[illegible]utes are used originally to form the[illegible] an item, and (*b*) the cue effective-ness of attributes is enhanced by their prior use[illegible]eval cues. Long-term recall per-form[illegible]ill be aided by an earlier recall to the[illegible] that the same attributes can be used[illegible]trieval cues on both tests.

REFERENCES

Atki[illegible], R. C., & Shiffrin, R. M. Human me[illegible] A proposed system and its control p[illegible] In K. W. Spence & J. T. Spence (Ed[illegible]), *The psychology of learning and motivation.* V[illegible] New York: Academic Press, 1968.

Atki[illegible] R. C., & Shiffrin, R. M. The control of [illegible] memory. *Scientific American*, 1971, 224[illegible], 82–90.

Coo[illegible] H., & Pantle, A. J. The total-time hy[illegible] in verbal learning. *Psychological Bul-let*[illegible] 68, 221–234.

Crai[illegible] M. The fate of primary memory item[illegible] free recall. *Journal of Verbal Learning an*[illegible] *Behavior*, 1970, 9, 143–148.

Craik[illegible] M., & Lockhart, R. S. Levels of pro[illegible] A framework for memory research. *Jo*[illegible]*erbal Learning and Verbal Behavior*, 1972, 11, 561–565.

Jacoby, L. L. Encoding processes, rehearsal, and recall requirements. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 302–310.

Jacoby, L. L., & Bartz, W. H. Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 561–565.

Lachman, R., & Mistler, J. L. Rehearsal, test trials, and component processes in free recall. *Journal of Experimental Psychology*, 1970, 85, 374–382.

Meunier, G. F., Ritz, D., & Meunier, J. A. Rehearsal of individual items in short-term memory. *Journal of Experimental Psychology*, 1972, 95, 465–467.

Müller, G. E. Zur Analyse der Gedächtnis-tätigkeit und des Vorstellungsverlaufes. *Zeit-schrift für Psychologie*, 1911, Suppl. 5.

Rundus, D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 1971, 89, 63–77.

Rundus, D., & Atkinson, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 99–105.

Thorndike, E. L., & Lorge, I. *The teacher's wordbook of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Pub-lications, 1944.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Underwood, B. J. Are we overloading memory? In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, D.C.: V. H. Winston, 1972.

(Received May 26, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 298–307

VARIED FUNCTIONS OF PUNISHMENT IN DIFFERENTIAL INSTRUMENTAL CONDITIONING[1]

GEORGE T. TAYLOR[2]

State University of New York College at Potsdam

A simple discrimination setting—differential instrumental conditioning—was used to assess the functions of punishment in discrimination learning in a 2 × 2 × 3 factorial design, plus 2 control groups, with 168 rats. The variables manipulated were type of stimulus (tone or shock), presentation of the trials (forced–free or forced), and alley of stimulus presentation (postive stimulus, negative stimulus, or both). The results indicated that shock can serve a variety of functions, e.g., as (*a*) an aversive stimulus, (*b*) a distinctive, informative stimulus, (*c*) a secondary reinforcer, and (*d*) perhaps most interestingly, an alerting stimulus. These findings are interpreted as suggesting that the role of punishment in differential instrumental conditioning is complex.

A response-produced aversive stimulus, i.e., punishment, does not always reduce the probability of a response. A notable example is a study reported by Muenzinger (1934) which demonstrated that, in a discrimination setting, shocking rats in the presence of the positive stimulus (S+) may facilitate learning the discrimination. This finding, the shock–right phenomenon, has generated a large body of data and theoretical speculations which relate to the more comprehensive properties of aversive stimulation.

Muenzinger (1934; Muenzinger & Wood, 1935) interpreted his data as implicating an effect of punishment which was general in nature. The function of noxious stimuli was suggested to be that of alerting the animal so as to make him respond more readily to the significant cues in the learning situation. The general alerting mechanism was proposed to be revealed behaviorally as vicarious trial-and-error (VTE) activity (Muenzinger, Bernstone, & Richards, 1938).

Later, Wischner (1947) questioned the generality of these results and subsequent conclusions, especially with respect to Muenzinger (1934; Muenzinger et al., 1938; Muenzinger & Wood, 1935) using only correction procedures in which the animal must retrace his incorrect responses. Using the noncorrection procedure, Wischner found that punishing the correct response appeared to only slightly facilitate discrimination efficacy relative to a nonshocked group, although the differences between the 2 conditions were small and statistically unreliable. However, though the difference was not as great as that found earlier by Muenzinger (1934), Wischner replicated the basic shock–right phenomenon. Thus, even in the noncorrection paradigm, punishment in the presence of S+ facilitated the discrimination rather than the expected inhibitory effect.

An alternative interpretation of these data emphasized the specific effects of punishment (cf. Fowler & Wischner, 1969). For example, the shock, by virtue of its association with the reward, served as a discriminative stimulus and, therefore, as a conditioned stimulus (Azrin & Holz, 1966). Extending the discriminative stimulus analysis, Fowler and Wischner (1965) have suggested that shock for either correct or incorrect responses served to differentiate the alleys more quickly and thus delimit the generalization between the 2 alternatives. As a result, these authors (Azrin & Holz, 1966; Fowler & Wischner, 1969) have re-

[1] This article is based upon a dissertation submitted to the University of New Mexico in partial fulfillment of the requirements for the PhD degree. The author would especially like to thank Frank A. Logan who served as the committee chairman. The research was performed while the author was a National Institute of Mental Health trainee and was supported in part by National Science Foundation Grant GB-12954 to Frank A. Logan.

[2] Requests for reprints should be sent to George T. Taylor, Psychology Department, State University of New York College at Potsdam, Potsdam, New York 13676.

jected the general alerting analysis of Muenzinger et al. (1938) and have adopted the position that the facilitative consequence of shock–right training is based upon shock serving as a secondary reinforcer and/or as a distinctive stimulus.

Although these 2 interpretations are not mutually exclusive, and indeed both may be implicated, theorists, in the interest of parsimony, have been reluctant to accept both to account for the shock–right data. It is possible that the resulting controversies (cf. Fowler & Wischner, 1969) stem from the extensive use of the simultaneous discrimination paradigm. This procedure has several disadvantages for an analysis of the effects of shock. First, it enables the animal to compare directly the discriminanda. Thus, if VTE, or some comparable general mechanism, contributes to the shock–right phenomenon, all *S*s had similar opportunities to make such comparisons. Second, and perhaps more importantly, the paradigm provides only one measure of performance, choice behavior. This measure represents the differences between the response strengths of the alternatives, but does not indicate directly which is being affected by the shock.

Accordingly, the present experiment attempted to make a unique contribution by including a simpler discrimination setting—differential instrumental conditioning—in which the discriminable cues are presented singly. This paradigm, therefore, provides a procedure which limits the opportunity to compare the stimuli directly, as well as providing an individual measure of response strength to each alternative. In addition, a more neutral tone stimulus was given to some *S*s to allow an assessment of the aversiveness of shock.

Method

Subjects. The *S*s were 168 female albino rats obtained from a commercial supplier. The experimentally naive *S*s were 90–120 days of age at the beginning of experimentation and were housed individually with water freely available. A daily ration of 12 gm. of laboratory chow was given immediately after each day's experimental session. The *S*s, randomly assigned to 1 of 14 groups with 12 *S*s in each, were run in replications of 2 *S*s per cell; each treatment was represented in each replication.

Apparatus. The apparatus was a trio of parallel alleys, each 2.44 m. long. The 2 outside alleys were black and the middle alley was white, thereby allowing spatiality of stimulus presentation as a nonrelevant cue on choice trials. On forced trials, the gray start box was positioned *directly* in front of the appropriate black or white alley and the adjacent alleys were blocked from *S*'s view. The free-choice trials, however, involved placing the start box in such a position that *S* could view both a black and a white alley. Thus, assurances were made so that *S* could view both discriminanda only in the free-choice setting. The apparatus and the start-box arrangement are described in detail by Logan (1960). Photocells were located 10 cm. from the start door, after the first 1.22 m., and directly before the food cup.

Grid floor inserts completely covering .61 m. of the alley and through which shock could be delivered were placed in each of the 3 alleys directly preceding the final .61 m. of the runway. A speaker connected to a tone generator and amplifier, located outside the apparatus, was positioned beneath the grid floor inserts. The start door opened vertically by a spring when released by a solenoid. Goal-box doors lowered manually enclosing the final .61 m. of the alley. Reward consisted of 45-mg. Noyes pellets delivered by Davis feeders.

Shock was delivered by 120-v. ac passing through a variac into a step-up transformer with 120-kΩ resistance in series with *S*. The grid floor consisted of .6-cm. stainless steel rods .3 cm. apart. The shock stimulus was 60 v. at .2-sec. duration, a value used frequently in the relevant literature (cf. Fowler & Wischner, 1969). The 1,000-Hz., 60-db. tone was presented at .2 sec.

Procedure. Pretraining began 1 wk. prior to the beginning of experimentation when *S*s were handled and placed on 23-hr. food deprivation. The *S*s then received 5 days of exploration, goal-box placements, and magazine training in the 3 alleys.

Brightness-discrimination training commenced with the introduction of differential instrumental conditioning. The *S* received 9 pellets in the S+ alley, black for half of the *S*s in each group and white for the remaining *S*s, and 1 pellet reward in the other (negative stimulus, S−) alley. Each *S* received a total of 8 trials/day for 20 days. For *S*s in one set of conditions (forced-free), the first 4 trials of each 8-trial block were forced according to one of the following rotated sequences: + + − −, − − + +, + − − +, − + + − (Logan, 1969). Trials 5 and 6 were free-choice trials and were randomly switched so that over a block of 8 choice trials, the positive alley appeared an equal number of times on the left and the right. Trials 7 and 8 were then forced opposite to those chosen on the free-choice trials (Logan, 1969) The *S*s in the other set of conditions (forced) received only differential conditioning trials in each 8-trial block. The first 4 trials were sequenced as above and the last 4 trials were opposite to the first trials' sequence.

A trial began by placing *S* into the start box and

releasing the start door. When broken, the first photobeam measured starting times on forced trials and choice times on the free trials. When the second photobeam halfway down the alley was broken, *S* received either a shock or a tone, if appropriate. Thus, for example, a shock–right *S* received a shock only when in the S+ alley. A shock–both or tone–both *S*, however, was presented with the appropriate cue on half of the trials in each alley, thereby equating the total number of stimulus presentations. These *S*s received the appropriate stimulus on one positive and one negative trial of the first 4 trials, randomly sequenced from day to day, and then on the opposite positive and negative trials of the second 4 trials of the day. The *S*s were run in squads of 4, providing an intertrial interval of 3–4 min.

Experimental design. A 2 × 2 × 3 factorial, plus 2 control groups, was the experimental design employed. The variables manipulated were type of stimulus (tone or shock), presentation of the trials (forced–free or forced), and alley of stimulus presentation (S+, S−, or S±). The 2 control groups received no stimulus, one group was given forced–free trials and the other group was given only forced trials.

Results

Since all *S*s received 4 initial differential conditioning trials each day, 2 in the positive alley and 2 in the negative alley, those trials formed the primary basis of comparison. The start speeds served as the major dependent variable, in part as a function of the data gen[illegible] in the setting and in part on theor[illegible] [illegible]ounds.

The shock was administered [illegible] middle of the alley. The data dem[illegible]te increased reductions in alley sp[illegible] as the shocked *S*s neared the shock lo[illegible]. These data are consistent with the hy[illegible]sized approach–avoidance gradients [illegible] Miller, 1944), i.e., the avoidance gradie[illegible] [illegible]mes increasingly steep, relative to the [illegible]ach gradients, as *S* nears the shock loca[illegible]. As a result, the *alley* speeds were con[illegible]ded by the presence of shock within th[illegible]ley, making comparisons between shock [illegible]ditions and between shock and no[illegible]ock conditions based on alley speeds diffi[illegible]lt, at best.

The *goal* speeds were confounded because shock served as an intense motivation[illegible]l variable (Spence, 1956) and shocked *S*s ran into the goal compartment unusually fast following presentation of the aversive stimulus. The start speeds were the least confounded measure and could better reflect differential performance of *S*s without the additional and confounding roles of shock, i.e., its aversive and motivation[illegible] capacities. Finally, since the start speed of *S*s with the black stimulus as the S[illegible]

TABLE 1

Absolute Mean Start Speeds (in Reciprocals) of Shocked Animals on Differential Conditioning Trials for Positive (S+) and Negative (S−) Alleys

Group	2-day blocks								
	2	3	4	5	6	7	8	9	10
Shock-right (FF)									
S+	1.13	1.11	1.28	1.13	1.33	1.28	1.19	1.21	[illegible]
S−	1.07	1.11	1.05	.82	.83	.75	.60	.48	.40
Shock-right (F)									
S+	1.10	1.33	1.23	1.17	1.25	1.29	1.18	1.19	[illegible]
S−	1.02	1.17	1.11	.93	1.10	.80	.73	.71	[illegible]
Shock-wrong (FF)									
S+	1.20	1.31	1.39	1.38	1.47	1.50	1.52	1.51	1.53
S−	1.17	.83	.65	.65	.52	.59	.58	.57	.55
Shock-wrong (F)									
S+	.98	1.29	1.19	1.45	1.35	1.50	1.37	1.37	1.41
S−	.93	1.08	.90	.85	.85	.83	.73	.70	.69
Shock-both (FF)									
S+	.83	.80	.81	1.03	1.12	1.28	1.29	1.24	1.27
S−	.73	.87	.94	.76	.62	.72	.50	.58	.57
Shock-both (F)									
S+	.67	.87	.91	1.06	1.04	1.18	1.11	1.10	1.12
S−	.73	.88	.86	.74	.74	.73	.65	.63	.62

Note. Abbreviations: F = forced trials, FF = forced-free trials.

TABLE 2

Absolute Mean Start Speeds (in reciprocals) for Tone Groups on Differential Conditioning Trials for Positive (S+) and Negative (S−) Alleys

Group	2-day blocks								
	2	3	4	5	6	7	8	9	10
Tone-right (FF)									
S+	.84	1.42	1.67	1.58	1.69	1.69	1.85	1.84	1.83
S−	.83	1.32	1.44	1.33	1.08	.95	1.02	.82	.85
Tone-right (F)									
S+	1.09	1.47	1.65	1.63	1.73	1.71	1.80	1.83	1.78
S−	1.12	1.43	1.38	1.33	1.02	.92	.83	.88	.91
Tone-wrong (FF)									
S+	.90	.98	1.28	1.62	1.78	1.63	1.56	1.49	1.57
S−	.78	.93	1.12	1.20	1.19	.90	.65	.61	.63
Tone-wrong (F)									
S+	.99	1.28	1.30	1.51	1.64	1.70	1.61	1.58	1.55
S−	.98	1.19	1.08	1.09	.99	.76	.80	.79	.77
Tone-both (FF)									
S+	1.13	1.22	1.19	1.31	1.27	1.42	1.40	1.39	1.41
S−	1.08	1.13	1.17	1.20	1.12	.88	.82	.78	.77
Tone-both (F)									
S+	.95	1.12	1.28	1.32	1.41	1.42	1.52	1.39	1.30
S−	.92	.99	1.09	1.20	1.23	1.00	.74	.63	.59

Note. Abbreviations: F = forced trials, FF = forced–free trials.

did not differ with their respective counterbalanced *Ss* with the white stimulus as the S+, those subgroups were combined for the subsequent analyses.

Absolute start speeds. The start speeds to S+ and S− were averaged and appear in tabular form (Tables 1–3). These data suggest that only in the shock groups did the forced–free procedure produce superior differential performance relative to the forced condition. Two analyses of variance of the S+ and S− start speeds over all trials, with all groups included, yielded significant values for both alleys, F (13, 154) = 2.25, $p < .05$, and F (13, 154) = 7.59, $p < .01$, respectively.

A posteriori comparisons (Scheffé method) indicated that the forced–free trained shock–wrong *Ss* ran more slowly in the S− alley than the other *Ss*. The other forced–free *Ss* receiving shock in the S− alley (shock–both) ran slower in that alley than the nonshocked *Ss*, with the notable exception of *Ss* with an informative tone stimulus at that locus, i.e., tone-wrong. Therefore, shock in the presence of the negative stimulus had a significant inhibiting effect, with respect to start speeds, in the S−

TABLE 3

Absolute Mean Start Speeds (in reciprocals) for No-Stimulus, Control Animals on Differential Conditioning Trials for Positive (S+) and Negative (S−) Alleys

Group	2-day blocks								
	2	3	4	5	6	7	8	9	10
No stimulus (FF)									
S+	.80	1.23	1.17	1.12	1.51	1.41	1.42	1.32	1.37
S−	.94	1.19	1.11	1.30	1.15	1.04	.74	.66	.70
No stimulus (F)									
S+	1.08	1.29	1.38	1.36	1.45	1.47	1.44	1.35	1.37
S−	1.13	1.35	1.29	.98	.91	.80	.78	.61	.70

Note. Abbreviations: F = forced trials, FF = forced–free trials.

alley. Subsequent comparisons of the groups' performance in the S+ alley indicated no 2 groups differed. One may note that there was the suggestion of differences; however, the error term was sufficiently large to yield no statistically reliable differences.

Relative start speeds. Because shocked *Ss* ran slower than nonshocked *Ss*, relative start speeds to the 2 discriminanda are more appropriate than the absolute speeds for statistical comparisons of discrimination performance. Two relative measures were employed. One such measure (Goldstein & Spence, 1963) is simply the difference between the S+ and S− start speeds. These difference scores are plotted in 2-day blocks in Figure 1. A 2 × 4 factorial analysis of variance was performed on the difference scores of the shock and no-stimulus groups (Figure 1A–1D) with type of training (forced–free or forced), and alley of stimulus presentation (S+, S−, *S*±, or neither S±), as the main factors. The results indicated that both training and stimulus variables significantly influenced responding, F (1, 88) = 31.13, $p < .01$, and F (3, 88) = 47.49, $p < .01$. A posteriori comparisons indicated that each shock group trained under a forced–free procedure yielded greater differences between S+ and S− speeds than its respective shock group receiving only forced trials. Also, shock–wrong groups learned the discrimination more rapidly than their comparably trained shock–right, shock–both, and no-stimulus groups, which did not differ.

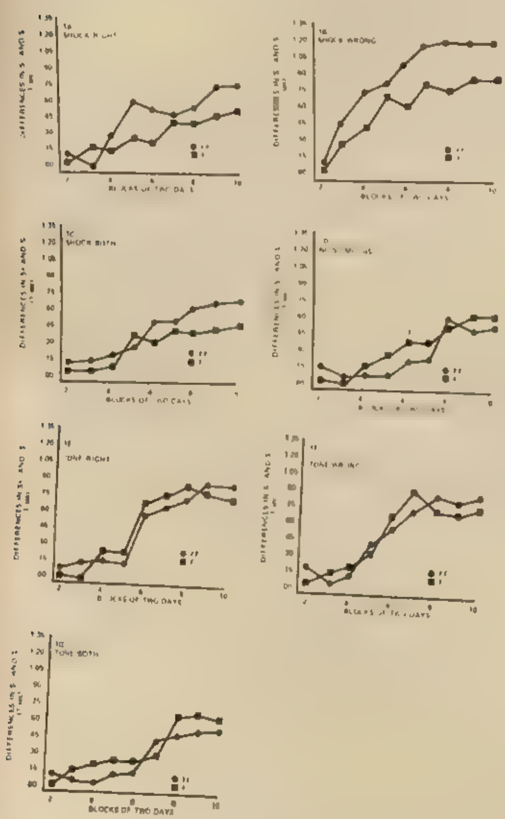


FIGURE 1. The difference between start speeds on the differential conditioning trials in the positive (S+) and negative (S−) alleys. (Each stimulus condition is plotted with its respective forced–free and forced groups.)

A separate 2×4 factorial analysis with the same factors was performed on the data of the tone and no-stimulus groups. As suggested in Figure 1D–1G, the different stimulus presentation procedures had no statistically reliable effect upon the various groups' performance, F (1, 88) < 1. However, the alley of stimulus presentation variable exerted significant effects upon discrimination efficiency, F (3, 88) = 121.01, $p < .01$. A posteriori analyses showed that the performance of tone groups with information (tone-wrong and tone right) was superior to the groups lacking an added informative cue (tone-both and no stimulus).

Thus, the start speeds indicate that interspersing free-choice trials into differential conditioning training produced greater differences between S+ and S− responding only for Ss presented with an aversive stimulus. Furthermore, a redundant, informative stimulus, either tone or shock, enhanced differential responding relative to a no-stimulus control condition.

Correct response measure of speeds—within stimulus conditions. A second way of measuring differential performance, which may be less affected by absolute levels, is the method suggested by Grice (1966). Successive blocks of differential conditioning trials (8 trials in the present experiment) were ranked from fastest to slowest and

ed with respect to the U statistic. rrect response was defined as S running aster in the S+ alley than in the alley. The percentage of correct re-ises for the different stimulus groups is tted in Figure 2. Again the figure gests that the shocked groups differ with spect to number of correct responses on he discrimination under the 2 types of training. A 2 × 4 factorial analysis on percentage of correct responses by the shock groups and the no-stimulus controls (Figure 2A–2D) revealed both type of training and stimulus presentation to be significant variables, F (1, 88) = 24.19, $p < .01$, and F (3, 88) = 38.33, $p < .01$. A posteriori comparisons indicated that each forced–free shock group made more correct responses than its respective forced shock group. Also the shock–wrong groups performed more efficiently than the comparably trained shock–right, shock–both, and no-stimulus groups. Finally, the shock–right and shock–both *S*s made similar numbers of correct responses and both were superior to the similarly trained, forced–free no-stimulus *S*s.

Next, a 2 × 4 factorial analysis was applied to the percentage of correct responses by the tone and no-stimulus *S*s. As suggested in Figure 2D–2G, alley of stimulus presentation yielded a significant value, F (3, 88) = 28.70, $p < .01$, which subsequent a posteriori analyses revealed to be a function of the information groups (tone–right and tone–wrong) making more correct responses than the noninformation groups. In contrast to the speed differences scores, the type of training was also a significant factor, F (1, 88) = 7.38, $p < .01$. However, a posteriori comparisons indicated that there were no differences between forced–free trained *S*s and their comparable but forced *S*s.

Therefore the graphical evidence that the type of training was a significant variable only for the shock groups and alley of stimulus presentation was a significant factor for both shock and nonshock conditions was confirmed statistically with both difference scores and the measure of correct responding.

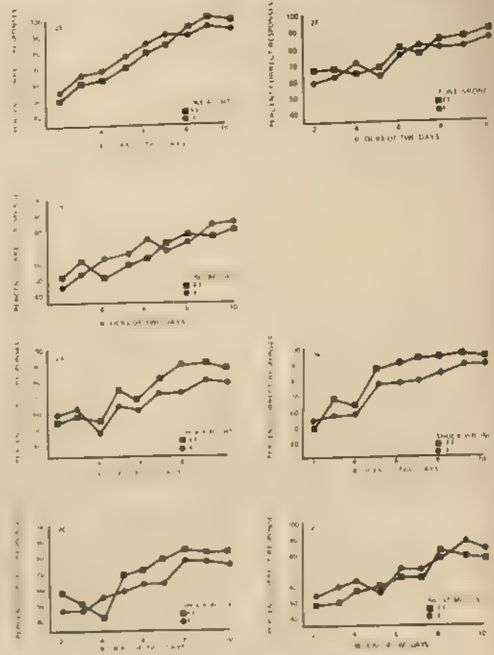


FIGURE 2. The percentage of correct responses on the first 4 differential conditioning trials in blocks of 2 days (see text for explanation). (Each stimulus presentation group is shown in relation to its respective training condition, either forced-free or forced.)

Correct response measure of speeds—between stimulus conditions. Since comparisons of shock, tone, and no-stimulus groups with comparable stimulus presentations were of interest, the correct response measure was employed. Three 2 × 3 analyses of variance with days as a repeated factor were performed with a major interest in the Stimulus Presentation × Days interaction. The assumption was that a significant interaction would suggest different learning rates by the *S*s, a finding relevant to a number of predictions.

Analysis of correct responses of the shock–right, tone–right, and no-stimulus control *S*s yielded significant differences between the various stimulus conditions, F (2, 66) = 17.38, $p < .01$. The Days × Stimulus interaction was also statistically reliable, F (16, 528) = 4.25, $p < .01$. Subsequent a posterior comparisons between the groups indicated that tone–right *S*s trained under both procedures and forced–free trained shock–right *S*s performed more

efficiently than no-stimulus and forced shock–right *S*s. Thus one may conclude that a stimulus in the correct alley facilitates learning the discrimination, with respect to fewer errors and faster rates of learning. However, when shock is used as the cue, this conclusion is valid only if *S*s are allowed free as well as forced trials.

A 2 × 3 factorial analysis was then performed, again with days as a repeated factor, for *S*s with the informative stimulus in the *S*– alley and no-stimulus *S*s. The results indicated that both the stimulus variable and the Days × Stimulus interaction yielded statistically reliable results, $F\ (2, 66) = 48.20$, $p < .01$, and $F\ (16, 528) = 6.40$, $p < .01$. A posteriori analyses between the groups indicated that shock–wrong groups were statistically superior to tone–wrong groups which, in turn, were more efficient than no-stimulus groups. Therefore, shock in the incorrect alley greatly hastened learning the discrimination, and a tone stimulus at that locus produced facilitation relative to the no-information control groups.

A final 2 × 3 factorial analysis with days as the repeated measure was performed on the correct responses of the noninformative groups (shock–both, tone–both, and no stimulus). Although no overall statistical differences were found between the various groups, $F\ (2, 66) < 1$, the Stimulus × Days interaction was a significant factor, $F\ (16, 528) = 2.25$, $p < .01$. A posteriori comparisons between the various conditions revealed that shock–both *S*s trained under the forced–free procedure made

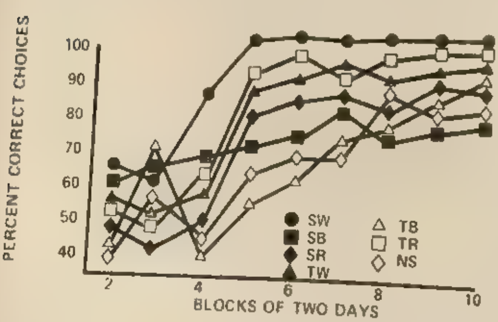


FIGURE 3. The percentage of correct choices on the free-choice trials blocked over 2 days for the forced–free groups.

significantly more correct respon· · than the other noninformative groups. There-fore these data suggest that an u· ·liable aversive stimulus enhanced discri· ·ation performance if *S*s were allowed fre· · well as forced trials.

The general conclusion of the dire· · comparisons of shocked *S*s with thei· ·one-stimulus counterparts and no-stimul· control *S*s indicates that shock in the in· ·rect alley and in both alleys improve· the differential conditioning preformance · ·er tone–wrong and tone–both trainin· ·e-spectively. This generalization is ·n-sonant with the data only insofar as ·he training of the shocked *S*s allows for f· ·e-choice trials. Finally, shock in the cor· ·t alley does *not* improve performance rela to tone–right training, though both to· right and shock–right facilitated perfor· ance relative to no stimulus.

Choice trials. The results of the choic· measure are plotted in Figure 3 alongside the results of the differential conditioning trials. The figure suggests similar ordering of the groups with respect to percentage of correct choices and percentage of correc· responses, with the possible exception tha the shock–right and shock–both perform· ance appears somewhat poorer on the fre· choice trials.

An analysis of variance on the percentag· of correct choices yielded a significant valu·, $F\ (6, 77) = 26.03$, $p < .01$. Subsequent a posteriori analyses indicated that the shoc· wrong group made significantly more correct choices than the other groups. Also, the 2 information groups (tone–right and tone–wrong) performed more efficiently than the no-stimulus control *S*s. Finally, the shock–right and shock–both, tone–both, and no-stimulus groups performed comparably on the free-choice trials.

Summary. Most of the results of the present study are clear and highly reliable. First, interspersing some free-choice trials into a predominantly differential conditioning procedure facilitated discrimination performance, but only for those conditions involving an aversive event. This generalization may be restricted to the relatively simple black–white discrimination

employed in the present investigation, but it is noteworthy. Second, providing an informative, albeit redundant cue in the negative alley facilitated discriminative behavior, and more so if that cue was aversive. Providing an informative, redundant cue in the positive alley was facilitative in the case of the neutral, tone stimulus. However, with respect to the aversive, shock stimulus in the S+ alley, only with one measure, percentage of faster speeds in S+ than in S−, were the effects positive. There was no difference either in raw speed differences or choice behavior. Finally, a noninformative cue produced improved discrimination performance, at least in one measure, but only if that cue was noxious.

Discussion

The results suggest that the role of punishment in differential instrumental conditioning is complex. There is extensive evidence in the data to support the notion that shock in a relatively simple discrimination setting can serve a variety of functions, e.g., as (*a*) an aversive stimulus, (*b*) an informative stimulus, (*c*) a secondary reinforcer, and (*d*) an alerting stimulus.

Consider first the aversiveness or negative incentive value of punishment. Logan (1969) has presented data which suggest that an animal is choosing continuously over time to determine the *net* incentive motivation value associated with an alternative, including not only the parameters of reward, but also the parameters of punishment. Thus, an animal weighs the negative and the positive incentive properties associated with the various response alternatives and responds on the basis of this net incentive value.

This role was clearly evident in several aspects of the present data. The shocked *S*s generally ran more slowly than their nonshocked counterparts. Furthermore, the S− speeds of *S*s receiving the aversive stimulus in the negative alley were reduced relative to the S− performance of *S*s shocked in the positive alley. Also, differential performance developed more rapidly for shock–wrong *S*s, i.e., *S*s with the greater difference in net incentive between the 2 response alternatives, than for any other condition. Finally, a shock stimulus in the positive alley was not as beneficial as a neutral, tone stimulus, suggesting that the differential in incentive between the alternatives was thereby reduced for the shock–right *S*s. Accordingly, this study, including differential conditioning, substantiates the earlier studies (e.g., Fowler & Wischner, 1965) employing simultaneous discrimination learning paradigms in indicating that one of the roles of shock is to contribute negative incentive selectively to the alternative in which it is presented.

A second hypothesized role of shock is its distinctive stimulus or information value, more specifically its secondary reinforcing value in the shock–right procedure. The present data are also generally consistent with this interpretation. The finding that an informative tone facilitated discrimination performance is strong evidence that a redundant, predictive cue is beneficial. An informative shock, as a distinctive stimulus, certainly could have served the same function. Also, that tone–wrong *S*s were as slow in the S− alley as shocked *S*s, with the exception of the shock–wrong condition, is further evidence of the cue value of a stimulus at that locus. Finally, the fact that, at least in one measure, the shock–right procedure was superior to the no-stimulus procedure further attests to this role, in spite of the negative incentive properties of the shock.

Interest then centers on whether there is yet a third role of shock in discrimination learning. Several authors (Muenzinger, 1934; Tolman, 1948) have proposed that shock serves an alerting function in making the animal more sensitive to the stimuli in the experimental setting. The unique contribution of the present study is employing differential conditioning to examine the properties of shock. The data of most interest, therefore, are the speeds on the differential conditioning trials. When the start speeds were transformed to a measure of correct responding, i.e., faster speeds in the positive than in the negative alley, to provide for comparisons of shocked and nonshocked *S*s, the data were highly supportive of an alerting role of aversive stimulation.

The finding that providing shocked *S*s free-choice trials facilitated differential responding, while not aiding the nonshocked *S*s, is consistent with a general alerting interpretation. Muenzinger et al. (1938) contend that shock produces increased amounts of VTE resulting in more efficient discrimination performance. Reducing the opportunity to VTE and comparing the stimuli by giving only single presentations of the stimuli should retard discrimination efficacy. A specific effects analysis, e.g., distinctive stimulus properties (Fowler &

Wischner, 1969), would make no such differential predictions with respect to training procedures. The present data, implicating a comparison of stimuli consequence of shock, support an alerting role of punishment such as that conceptualized by Muenzinger et al.

An important prediction from an alerting analysis is that shock, by virtue of its sensitizing effects, facilitates learning rates relative to comparably trained but nonshocked animals. This prediction was upheld by the data of shock–wrong and shock–both *S*s who had been given free-choice trials, for those *S*s learned the discrimination more rapidly than the similarly trained tone–wrong and tone–both groups. However, a shock or tone stimulus in the correct alley produced comparable learning rates. One may contend that specific effects, i.e., inhibition as a function of aversiveness, of the shock contributed to the shock–wrong facilitation and the lack of facilitation in the shock-right condition relative to their respective tone groups. As a result the shock–both findings take on even greater importance. Providing a nonspecific stimulus, i.e., not associated reliably with either the positive or negative alternative, hastens the learning rates, but only if that stimulus was noxious. Thus, these data also implicate a general alerting role of punishment in discrimination learning.

Of several theoretical formulations of a general effect of aversive stimulation, none appear entirely satisfactory. Tolman (1948) mentioned, but only briefly, that shock produces "looking around" by the animal to see what it was that had hit him. Muenzinger (1934; Muenzinger et al., 1938) proposed an alerting concept, but one which is totally dependent upon VTE behavior. Although the present results implicate a comparison factor, other investigators (e.g., Farlie, 1937; Wischner, 1947), and the present *E* through informal observation, found no reliable relationship between VTE and discrimination efficacy. A more realistic proposal is that the anticipation of shock slows the animal (Logan & Wagner, 1965) allowing it to vacillate at the choice point. This vacillation need not be overt, as is VTE but could be a changing covert perceptual orientation (Spence, 1936) which has an effect similar to that proposed earlier for VTE (Muenzinger et al., 1938), i.e., *S*'s increased sensitivity to the discriminanda.

In conclusion, the novel contribution of the present study involves examining the function of shock in a simple discrimination setting—differential instrumental conditioning. Analysis of the individual response tendencies to each stimulus, a measure not available in the simultaneous discrimination paradigm, indicated that shock serves an aversive role which reduces response strength to the punished alternative. Furthermore, examination of the behavior of animals with an informative, but neutral tone stimulus, a comparison not made previously, indicated that such a stimulus enhances discrimination performance. The implication is clear that the shock, as a distinctive stimulus, is serving a similar function. Finally, a comparison of simultaneous and single presentations of the stimuli, again not attempted previously, suggested a general alerting role of shock. Thus, the present experiment added several unique conditions to the usual shock–right paradigm and substantiated the earlier speculations that shock can serve a variety of roles in discrimination learning.

REFERENCES

Azrin, H. N., & Holz, W. C. Punishment. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application.* New York: Appleton-Century-Crofts, 1966.

Farlie, C. W., Jr. The effect of shock at the 'moment of choice' on the formation of a visual discrimination habit. *Journal of Experimental Psychology*, 1937, 21, 662–669.

Fowler, H., & Wischner, G. J. Discrimination performance as affected by problem difficulty and shock for either correct or incorrect responses. *Journal of Experimental Psychology*, 1965, 69, 413–418.

Fowler, H., & Wischner, G. J. The varied functions of punishment in discrimination learning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

Goldstein, A., & Spence, K. W. Performance in differential conditioning as a function of variation of magnitude of reward. *Journal of Experimental Psychology*, 1963, 65, 86–93.

Grice, G. R. Dependence of empirical laws upon the source of experimental variation. *Psychological Bulletin*, 1966, 66, 488–498.

Logan, F. A. *Incentive.* New Haven: Yale University Press, 1960.

Logan, F. A. The negative incentive value of punishment. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

Logan, F. A., & Wagner, A. R. *Reward and punishment.* Boston: Allyn & Bacon, 1965.

Miller, N. E. Experimental studies of conflict. In J. M. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.

Muenzinger, K. F. Motivation in learning: I. Electric shock for correct responses in the visual

discrimination habit. *Journal of Comparative Psychology*, 1934, **17**, 267–277.

MUENZINGER, K. G., BERNSTONE, A. H., & RICHARDS, L. Motivation in learning: VIII. Equivalent amounts of electric shock for right and wrong responses in a visual discrimination habit. *Journal of Comparative Psychology*, 1938, **26**, 177–186.

MUENZINGER, K. F., & WOOD, A. Motivation in learning: IV. The function of punishment as determined by its temporal relation to the act of choice in the visual discrimination habit. *Journal of Comparative Psychology*, 1935, **20**, 95–106.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427–449.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale University Press, 1956.

TOLMAN, E. C. Cognitive maps in rats and men. *Psychological Review*, 1948, **55**, 189–208.

WISCHNER, G. J. The effect of punishment on discrimination learning in a noncorrection situation. *Journal of Experimental Psychology*, 1947, **37**, 271–284.

WISCHNER, G. J., HALL, R. C., & FOWLER, H. Discrimination learning under various combinations of food and shock for "correct" and "incorrect" responses. *Journal of Experimental Psychology*, 1964, **67**, 48–51.

(Received May 29, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 308–313

MEANINGFULNESS, PERCEPTUAL GROUPING, AND ORGANIZATION IN RECOGNITION MEMORY[1]

HENRY C. ELLIS,[2] FREDERICK J. PARENTE, AND E. CHANDLER SHUMATE

University of New Mexico

Two experiments examined the role of meaningfulness (*m*), perceptual grouping, and organizational factors in recognition memory of consonant-vowel–consonant (CVC) trigram stimuli. Experiment I demonstrated that recognition memory of trigram doublets of the integrated form CVC-CVC is a direct function of the *m* value of the stimuli. When the same [illegible] *m* doublets were fragmented by physically separating the letters according to grouping rules, thus disturbing the initial unitary perceptual organization of the stimuli, the effects of *m* were eliminated. The second experiment demonstrated that the fragmented high *m* stimuli operated as if they [illegible] low-*m* stimuli in an associative learning task; their recognition was facilitated by training in which the stimuli were associated with representative [illegible] providing evidence for response-dependent stimulus encoding. In general, contextual variables such as verbal labels are increasingly important in enabling the encoding of stimuli that are low in meaningfulness, perceptual unity, or other indexes of codability.

The fractionated vs. integrated nature of stimuli is thought to be related to the ease with which they can be placed in memory. Highly integrated stimuli (e.g., stimuli high in meaningfulness, *m*) may be encoded quite rapidly as they are processed as a single unit, whereas fractionated stimuli (e.g., low-*m* stimuli) are encoded more slowly because the encoding consists of more than 1 component (Shepard, 1963). This particular account of the underlying dimensions of *m* bears a strong resemblance to Miller's (1956) concept of "chunking," wherein units that may be integrated into larger chunks (and therefore fewer discrete ones) are more easily processed. In a similar vein, variations in stimulus *m* may produce corresponding demands upon *S*'s memory load, with the *m* value of stimuli being negatively correlated with memory load (Ellis, 1973). If low-*m* stimuli are regarded as more difficult to process because of their fragmentary, unintegrated nature, it should be possible to alter the *m* value of a unitary or high-*m* stimulus, and thus the ease with which it can be encoded as a unit, by varying the perceptual grouping of a sequence of letters normally perceived and encoded as a unit.

The primary purpose of the present experiments is threefold: (*a*) to demonstrate a positive relationship between stimulus *m* and recognition, (*b*) to determine if the effect of stimulus *m* depends upon the degree of integration of the stimuli, or the ease with which they may be perceptually grouped into a unit, and (*c*) to determine if fragmented high-*m* stimuli operate as low-*m* stimuli in the sense that their recognition is, to a great extent, dependent upon contextual cues present during associative training.

The first experiment examined the effect of stimulus *m* on recognition memory with both integrated and fractionated letter sequences. The recognition memory test provides a sensitive measure of the extent to which *S* has encoded the stimulus in a unitary fashion, because the test requires recognition of a target stimulus compared with a distractor item high in formal similarity to the target item. Although some studies (e.g., Ellis & Shumate, 1973) have demonstrated a positive relationship between stimulus *m* and recognition, negative relationships have also been reported (e.g., McNulty, 1965); moreover, the rela-

[1] This research was supported by National Science Foundation Grant GB-27413X to the first author, and by a National Defense Education Act graduate fellowship to the third author.

[2] Requests for reprints should be sent to Henry C. Ellis, Department of Psychology, University of New Mexico, Albuquerque, New Mexico 87131.

tionship also known to depend upon the scale of m [illegible]. Ellis, Muller, & Tosti, 1966). Thus, it was deemed necessary to establish this relationship for purposes of obtaining baseline data. The experimental question was: Does the effect of stimulus m (as defined by conventional norms) on recognition memory depend upon the integrated vs. fractionated character of the letter sequences?

The second experiment provided evidence for the view that fractionated high-m stimuli do, in fact, behave as low-m stimuli in an associative learning task. Several experiments have shown that recognition memory for visual form stimuli (Ellis & Daniel, 1971, Experiments I & II) and verbal materials (Ellis & Shumate, 1973; Ellis & Tatum, 1973) may be enhanced as a result of associative training. When stimuli are paired with "representative" verbal responses (labels), the stimuli are subsequently better recognized than when they are paired with "nonrepresentative" responses or are simply observed without associative training. Ellis (1973) has proposed that the facilitative effect upon recognition memory for stimuli paired with representative responses is dependent upon the initial codability of the stimuli, with relatively greater facilitative effects occurring with less codable stimuli. This proposition asserts that ambiguous, unintegrated, or complex stimuli are more dependent upon contextual events for their encoding than are familiar, integrated, or simple stimuli. Ellis and Shumate (1973), viewing m in terms of codability, with low-m stimuli being less codable than high-m stimuli, found that recognition memory facilitation for trigram stimuli paired with representative responses was dependent upon the m value of the stimuli, with reliable facilitation in recognition performance occurring with low-m stimuli, whereas no differences obtained with high-m stimuli.

The question of interest in the second experiment was whether associative training with the degraded high-m stimuli, using representative response terms, would produce an effect upon stimulus recognition memory similar to the findings obtained with low-m stimuli by Ellis and Shumate (1973). The experimental question was: Can a response which is representative of the stimulus provide, during associative training, cues that enable S to organize or integrate nominally fragmented stimuli? The assumption was that a representative response can effectively instruct S in how to organize or integrate a fragmented stimulus so that it is more readily encoded as a unit.

Experiment I

Method

Experimental design. The *Ss* were 40 undergraduate student volunteers from the University of New Mexico. Each S received 4 study–recognition test trials to learn a mixed list of 12 trigram doublets containing 3 levels of m. The *Ss* were randomly assigned to 1 of 2 list conditions in which the doublets were either "integrated," i.e., appeared as intact trigram pairs, or "fractionated" (degraded), i.e., broken up into particular letter groupings other than the original trigrams. The *Ss* assigned to the fractionated stimulus condition were tested 1 wk. later than those in the integrated stimulus condition, although both were drawn from the same population of *Ss*.

Stimuli. The integrated items were trigram doublets constructed by combining 2 trigrams selected from the Archer (1960) norms. For example, the trigrams BAN and COW were combined to form the doublet BANCOW. Doublets rather than trigrams were used in order to require S to encode more information and to allow for construction of different distractors on each recognition trial in accord with a single construction rule. Four doublets of each of the m levels (high, moderate, and low) were used, with average association values of 100%, 56%, and 12%, respectively. The doublets used were BANCOW, RUMWIG, CUPNET, DENGIT, PUZWEK, [illegible], [illegible], CIMBUX, NEJWOQ, TIJFRQ, YIWFEP, and MOJZEX, and were typed in [illegible]

The same doublets were used in the fractionated condition, in which each appeared as perceptual groupings of 1 or more letters. (If it is true that high-m stimuli are more easily encoded and hence better recognized than low-m stimuli because they can be more readily processed as a unit, then fractionation of the doublets should make the high- and moderate-m items more difficult to encode. Accordingly, recognition performance differences typically associated with variations in stimulus m should be either reduced or eliminated.) The 6 letters of each doublet were grouped according to 1 of 4 fractionation rules that simply regulated the number of letters in each group. For example, the 4 high-m doublets, BANCOW, CUPNET, RUMWIG, and DENGIT, were presented in the following forms, each

adhering to a different fractionation rule: BA NC OW, C UPN ET, R UMWI G, and DE NGI T. The moderate- and low-*m* doublets were also fractionated according to the same grouping rules, and each stimulus maintained the same form or grouping rule throughout the study–test sequence.

Procedure. All *Ss* were given 4 study–test trials to learn either list. The order of items was randomly varied on each trial. The doublets were presented consecutively on a Stowe memory drum at a 1-sec. presentation rate. The *Ss* were instructed to remember as many items as possible. After viewing all 12 items, *Ss* were given a 2-alternative forced choice recognition test in which they were shown each of the *old* items, singly, paired with a highly similar distraction item. Distractors were formally similar to the *old* items, with a distractor consisting of a 1-letter substitution of 1 of the 4 middle letters of the original doublet. A vowel was replaced by a vowel and a consonant was replaced by a consonant.

For *Ss* learning integrated items, the *old* item BANCOW appeared, for example, with the distractor BENCOW on one trial, BAZCOW on another, and so on. As *Ss* were presented with each *old–new* pair, they were instructed to underline the stimulus they had seen during the study trial. A different distractor item was used on each of the 4 trials, and the right–left position of the pairs was randomly varied on each trial. For *Ss* learning fractionated items, the same construction rules were used; however, the distractor items were fractionated according to the same grouping structure as the *old* items. For example, BA NC OW was paired with BE NC OW on 1 trial, BA ZC OW on another, and so on.

Results

Figure 1 is a plot of the proportion of correct recognitions of the trigram doublets for the conditions of the experiment. The left panel shows recognition performance curves for the integrated stimuli, and the right panel portrays the results for the fractionated stimuli. The left panel clearly shows the differential effects of *m* upon recognition learning, with high-*m* stimuli producing superior recognition performance, followed by moderate-*m* stimuli and low-*m* stimuli showing the poorest performance. In contrast, the right panel indicates that *m* produces no differential effect upon recognition when the items are fractionated. The mean proportions of correct recognitions of the integrated items, averaged over 4 trials for the high-, moderate-, and low-*m* conditions, were .86, .81, and .70, respectively. In contrast, for the fractionated items, the recognition scores averaged .62, .60, and .60, respectively. The interaction between *m* and stimulus grouping was reliable, $F\ (2, 76) = 4.51$, $p < .025$, which is the finding of principal interest. The effects of grouping (integrated vs. fractionated stimuli) was also highly reliable, $F\ (1, 38) = 41.03$, $p < .001$, and the overall effect of *m* was reliable, $F\ (2, 76) = 6.86$, $p < .01$. Separate analyses of the effect of *m* for the integrated stimuli yielded a reliable difference, $F\ (2, 38) = 11.14$, $p < .001$, whereas the effect of *m* with fractionated stimuli was not significant, $F\ (2, 38) < 1$, $p > .05$.

The figure shows the obvious effect of trials, $F\ (3, 114) = 85.00$, $p < .01$. The *Ss* learned the integrated items faster than the fractionated items, as reflected in the Trials × Grouping interaction, $F\ (3, 114) = 87.50$, $p < .01$, and the $m \times$ Trials interaction was reliable, $F\ (6, 228) = 5.00$, $p < .01$.

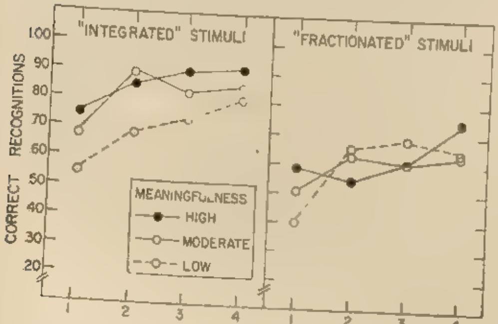


FIGURE 1. Recognition learning as a function of stimulus meaningfulness with integrated and fractionated trigrams.

EXPERIMENT II

The results of Experiment I provide direct support for the view that the effect of stimulus *m* on recognition performance is dependent upon the structural properties of the trigram doublets. High-*m* stimuli, when fragmented or grouped so that their unitary property is less readily detected, are recognized at approximately the same level as moderate- and low-*m* fragmented stimuli. It follows that if the high-*m* stimuli have indeed become functionally equivalent to low-*m* stimuli, then they should operate as low-*m* stimuli under other experimental conditions. Since it is known that low-*m*

stimuli are most susceptible to response-dependent encoding (Ellis, 1973; Ellis & Shumate, 1973), it would follow that degraded or fractionated high-*m* stimuli should show a similar effect. Ellis and Shumate (1973) have shown that the formation of an association between stimuli and responses can affect the recognizability of stimuli, and that this facilitative effect is evident only with low-*m* stimuli. This effect occurs with responses that are representative of the stimuli, but not with unrepresentative responses. In general, the conditions under which contextual cues aid in stimulus encoding are those with stimuli that are initially difficult to encode (Ellis, 1973). Therefore, it would follow that if fragmented high-*m* stimuli are paired with representative responses during associative training, such training would facilitate stimulus encoding compared with associative training in which the responses were nonrepresentative.

Method

The *Ss* were 70 undergraduate student volunteers from the University of New Mexico. Twenty *Ss* were assigned at random to each of 2 experimental conditions and 15 *Ss* to each of the remaining 2 conditions. All *Ss* were given 8 paired-associate (PA) trials to learn a list of 8 stimulus–response pairs. The stimulus terms were either integrated or fragmented high-*m* trigram doublets and were identical for both conditions except, of course, for the groupings. (The *Ss* assigned to the integrated stimulus conditions were tested 12 wk. later than those in the fractionated stimulus conditions. Nevertheless, *Ss* in both conditions were drawn from the same population of introductory psychology classes. Furthermore, the integrated stimulus conditions represent a replication of the immediate recognition condition of Ellis and Shumate (1973), and since quite comparable recognition results were obtained in that and the present experiment, the comparability of the groups is reasonably assured.) For 1 group of *Ss* the doublets were each paired with a representative response, and for the other group each doublet was paired with a nonrepresentative response. Each stimulus was exposed for a 4-sec. period consisting of a 2-sec. anticipation interval and a 2-sec. simultaneous presentation with the response. The intertrial interval was 4 sec. The *Ss* were tested individually, and the materials were presented via a Stowe memory drum. The response terms were common words, equated for frequency of occurrence between groups in accordance with the Thorndike–Lorge (1944) norms. Response representativeness (or belongingness) was determined by agreement among 6 judges as to whether or not a particular word selected "resembled" its stimulus mate. Judges were asked to rate each response as being representative in the sense of being descriptive, suggestive, or relevant to the stimulus. A nonrepresentative response was judged as bearing little, if any, relationship to its corresponding stimulus mate. These procedures are comparable to those employed by Ellis (1968) in studies of visual shape recognition. In the case of verbal stimuli, these operations basically mean selection of response terms that are formally and/or acoustically similar to the stimuli. Although the defining operations of response representativeness clearly imply a multidimensional concept, all that is necessary for the present experimental purposes is the selection of 2 classes of pairs for associative learning: 1 set of pairs that is relatively easy to associate and 1 that is relatively difficult. An example of pair types is M EDVA N associated with MAIDEN as a representative response, and with UMPIRE as a nonrepresentative response.

After *Ss* received PA training they were given a forced choice stimulus recognition test that required them to select 1 of 5 highly similar stimuli as being the *old* item seen during training. The 5 stimuli (the *old* item and 4 distractors) were typed in a single row on a white 5 × 8 in. card. The distractors were constructed as outlined in Experiment I and were grouped in the same manner as the *old* item with which they appeared.

Results

The primary data were the scores from the recognition test. Response representativeness exerted no effect on stimulus recognition with integrated high-*m* stimuli; mean recognition scores following either type of associative training were identical (.83). In contrast, with fractionated high-*m* stimuli, the recognition scores for the representative and nonrepresentative response conditions were .79 and .44, respectively. The expected Stimulus Structure × Response Representativeness interaction was present, F (1, 66) = 14.75, $p < .01$. Clearly, context effects during associative training are present, but only with the fractionated stimuli, which operate as poorly integrated units and thus benefit from contextual responses which effectively instruct *S* in how to unitize or organize the letter sequences.

The effects of stimulus structure (integrated vs. fractionated) and response representativeness were both reliable, F (1,

66) = 27.66, $p < .001$, and F (1, 66) = 12.60, $p < .001$, respectively.

Discussion

Experiment I provided baseline data demonstrating the effects of *m* on recognition memory for trigram doublet stimuli. The results indicated a positive and systematic relationship between the recognition of a stimulus and its *m* value. This finding is consistent with those of Ellis and Shumate (1973) using similar stimuli and with experiments using random shape stimuli (e.g., Ellis, 1968; Price & Slive, 1970). The first experiment provided empirical support for Shepard's (1963) hypothesis that *m* is highly correlated with the degree of analyzability of the stimuli. This hypothesis contends that low-*m* stimuli are likely to be analyzed into separate components, i.e., they are likely to be processed or encoded into more than 1 unit, whereas high-*m* stimuli are likely to be processed in a more unitary manner. Fractionation of high- and moderate-*m* stimuli produced a substantial reduction in recognition performance.

A straightforward interpretation of these findings is that the fragmented high-*m* stimuli have become, in effect, functional low-*m* stimuli. This position is reasonable if the view that the underlying dimension of *m* is the degree to which a verbal stimulus may be encoded as a unit is adopted. Fractionating a stimulus, particularly a high-*m* stimulus, makes it more difficult for *S* to encode that stimulus as a unit and therefore effectively reduces its *m* value. Relating these findings to Miller's (1956) chunking notions, a further interpretation of the results seems quite simple and straightforward. High-*m* stimuli may be regarded as involving fewer chunks of information than low-*m* stimuli, even though the number of letters in each stimulus class is identical. Thus, *S*s learning low-*m* stimuli attempt to place more units or chunks of information into memory than *S*s learning high-*m* lists. This differential memory loading, attributable to varying levels of *m*, is reflected in recognition memory performance, wi[illegible] learn[illegible] high-*m* material performing [illegible] than *S*s learning low-*m* material. [illegible], wh[illegible] high-*m* stimuli are fractionat[illegible] numbe[illegible] of chunks of information req[illegible] for h[illegible] learning to take place is inc[illegible] hereb[illegible] increasing *S*'s memory load[illegible] consi[illegible] quently lowering recognition [illegible]

The purpose of Experiment [illegible] to demonstrate that when a high-[illegible] stimul[illegible] is fractionated it does, in fact, [illegible] a low-*m* stimulus under different [illegible] mental conditions. The positive r[illegible] of the experiment suggest, again, th[illegible] the *m* effect is a function of the integ[illegible] fractionated nature of the stimulus. The consequence of fractionating a high-*m* stimulus is to make it more difficult for *S* to perceive and encode that stimulus as a unit. The function of a representative [illegible] sponse may be seen as one of reinteg[illegible] or organizing the stimulus by sugge[illegible] a possible unitary encoding. More ge[illegible] ally, this experiment suggests that [illegible] cedures that enable or encourage *S* to encode stimuli as units may be regarded as increasing the *m* value of that stimulus. By the same token, procedures th[illegible] courage *S*s to encode stimuli in [illegible] of their components may be seen as decreasing the *m* value of that stimulus. Finally, the results of Experiment II support the generalization that contextual events (such as responses) are most likely to affect the encoding of stimuli that are initially difficult to encode.

References

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, 74(10, Whole No. 497).

Ellis, H. C. Transfer of stimulus predifferentiati[illegible] to shape recognition and identification learning: Role of properties of verbal labels. *Journal* [illegible] *Experimental Psychology*, 1968, 78, 401–409.

Ellis, H. C. Stimulus encoding processes in hum[illegible] learning and memory. In G. H. Bower (Ed.) *The psychology of learning and motivation*. Vol. [illegible] New York: Academic Press, 1973.

Ellis, H. C., & Daniel, T. C. Verbal processes in [illegible] *Journal of Experimental Psychology*, 1971, 90, 18–26.

Ellis, H. C., Muller, D. G., & Tosti, D. T. Stimulus meaning and complexity as factors in

the transfer of stimulus predifferentiation. *Journal of Experimental Psychology*, 1966, 71, [illegible]
Ellis, H. C., & Shumate, E. C. Encoding effects of [illegible] and [illegible] on [illegible] memory of [illegible]. *Journal of Experimental Psychology*, 1973, 98, 70–[illegible]
[illegible] & Tatum, B. C. Stimulus encoding and the relationship between stimulus recognition and association formation. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 174–184.
McN[illegible] Short-term retention [illegible] of [illegible], recoding, and meaningfulness of the material. *Canadian Journal of Psychology*, 1965, 19, 188–196.
M[illegible] The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81–96.
P[illegible], R. H., & [illegible], A. B. [illegible] *Journal of Experimental Psychology*, 1970, 83, [illegible]
[illegible], R. N. [illegible] In [illegible] New York: McGraw-Hill, 1963.
Th[illegible] [illegible] New York: [illegible] Bureau of Publications, 1944.

(Received [illegible])

Manuscripts Accepted for Publication in the

Journal of Experimental Psychology

Evidence [illegible] Interference in Recognition [illegible]. Charles N. [illegible] (Department of Psychology, Pennsylvania State University, 417 Bruce V. Moore Building, University Park, [illegible] 16802), Andrew J. [illegible] and [illegible]

Attention and the Relation of Perception [illegible] (Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland).

[illegible] and [illegible] (Department of Psychology, Emory University, Atlanta, Georgia 30322) [illegible] and Philip S. [illegible]

Short-Term Retention of Visual Sequences [illegible] A. [illegible] (Department of Psychology, University of Nebraska at Omaha, Omaha, Nebraska 68101) and John G. Miscik.

Memory [illegible] and Long-Term Retention. Thomas O. N[illegible] (Department of Psychology, University of Washington, Seattle, [illegible]

[illegible] and [illegible] (Department of Psychology, State University of New York at [illegible]) and E. Webb Stacy, Jr.

Retrieval of Words from [illegible] Jonathan L. Freedman (Department of Psychology, Columbia University, New York, New York 10027) [illegible]

Relation between Similarity Grouping and [illegible] (Program, National Science Foundation, Washington, D.C. [illegible])

Partial Category Cuing: The [illegible] (University of [illegible], Oklahoma 73190) and Linda Warren.

Multidimensional Analysis of Choice Reaction Time [illegible] Pairs of English Fricatives. Frederick [illegible] (Department of [illegible], Hearing, and Speech Sciences, University of [illegible]) and Sadanand Singh.

The Effects of Segmentation and Expectancy on the Matching Time for Words and Nonwords. Arnold D. Well (Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01002), Robert M. Schindler and Alexander Pollatsek.

[illegible] Encoding and Decision Processes in Recognition Memory. James F. Juola (Department of Psychology, University of Kansas, Lawrence, Kansas [illegible]) [illegible]

The Role of Images within the Memory System. [illegible] (Department of Psychology, University of Maryland Baltimore County, 5401 Wilkens Avenue, Baltimore, Maryland 21228).

An Analysis of Recognition Performance [illegible] (Department of Psychology, Rice University, Houston, Texas 77001) and [illegible]

[illegible] Encoding and Recognition Memory [illegible] (Department of Psychology, Emory University, Atlanta, Georgia [illegible])

(*Continued on page 320*)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 314–320

CONNOTATIVE EVALUATION AND CONCRETENESS IN SHORT-TERM MEMORY[1]

GEORGE D. GOEDEL[2]

State University of New York College at Geneseo

Encoding processes were investigated using the release from proactive inhibition (PI) paradigm with word triads derived from the factorial manipulation of evaluative connotation (E) and concreteness (C). The recall data indicated significant PI buildup for all E×C conditions with less PI for concrete vs. abstract triads while no differential PI for good vs. bad triads. Proactive inhibition release was found for shifts on either E or C, or both, compared to their appropriate no-shift controls only when postshift items were abstract. For postshift abstract items, simultaneous shifts on E and C produced better recall than shifts on C only, but not better than shifts on E only. The data were evaluated in terms of a dual encoding mechanism for concrete word items and Paivio and Begg's offsetting interference hypothesis.

Considerable evidence has accumulated over the last decade for proactive inhibition (PI) and PI release with successive trials on a Brown-Peterson (Brown, 1958; Peterson & Peterson, 1959) short-term memory (STM) task. The use of PI release as a potential indicator of how information is encoded in STM has perhaps received its most cogent explication in a recent review by Wickens (1970). He proposed that words are encoded into memory along a number of dimensions or attributes and identified a large number of these dimensions, via PI release, gauging their relative dominance in terms of percentage of release.

The concern of the present study was with 2 such dimensions, connotative evaluation (E) as defined by the semantic differential (Osgood, Suci, & Tannenbaum, 1957) and concreteness (C). The fact that the former is employed as an encoding class in STM has been fairly well established (Turvey, 1968; Wickens & Clark, 1968). Evidence suggesting such encoding for the latter, however, has been lacking. Wickens and Engle (1970) [illegible] differential encoding by means of PI [illegible] for words differing on this dimension [illegible] although their results suggested differential interference for abstract vs. concrete word triads, their negative release results raised doubts concerning differential encoding.

The differential interference interpretation proposed by these authors, in terms of greater semantic overlap for abstract items, was recently challenged by Paivio and Begg (1971) who failed to find support for such an interpretation when interference was defined in terms of interitem associative relatedness. Paivio and Begg proposed instead that imagery is a source of resistance to interference among concrete items based on the assumption that high-imagery words are encoded differently from abstract words. According to dual encoding theory (Paivio, 1971), the availability of imagery as an alternative or additional code for concrete items might enhance their recall relative to abstract items by offsetting the interference effects of a verbal code, the only code available for abstract items. Unfortunately, Paivio and Begg did not include a shift in triad class in their study in which release could be evaluated. Thus, the Wickens and Engle (1970) failure to obtain release with this dimension has remained a question.

---

[1] This article is based in part on a dissertation submitted to Southern Illinois University at Carbondale in partial fulfillment of the requirements for the PhD degree. The author is indebted to Robert C. Radtke for his direction and guidance as chairman of the dissertation committee.

[2] Requests for reprints should be sent to George D. Goedel, Department of Psychology, State University of New York College at Geneseo, Geneseo, New York 14454.

The present study sought to explore further the phenomena of PI buildup over trials and PI release with concrete and abstract word triads by also including the connotative evaluative dimension. By using words rated on both E and C, shifts on one or both of these dimensions could be evaluated. The inclusion of the E dimension would provide a reliable encoding class under either an abstract or concrete triad condition and shifts on this dimension should result in PI release regardless of C level. This dimension was also chosen since it was not controlled in the Wickens and Engle (1970) study and there was the possibility that it may have been confounded somewhat with their imagery-concreteness manipulation. This possibility arose from the fact that these authors selected word stimuli from the Paivio, Yuille, and Madigan (1968) norms allowing imagery and C ratings to covary since they tend to correlate highly (.83). However, since Paivio et al. noted that words which differ on these 2 scales with imagery ratings being higher than C ratings tended to be words with "strong emotional or evaluative connotations," failure to control for E may have resulted in some confounding.

The present study also investigated the release phenomenon as a function of the number of prior PI buildup trials, since shifts in triad class occurred after either 4 or 8 prior interference trials. This variable was included since it was thought that perhaps recall performance had not reached asymptotic level prior to a shift in class in the Wickens and Engle study where only 3 preshift trials were given. This assumption appeared reasonable in the light of an earlier study by Wickens, Born, and Allen (1963) where significant release from PI was obtained with a shift from number–number–number to consonant–consonant–consonant trigrams relative to a consonant–consonant–consonant control after 9 prior PI buildup trials but not after 3 or 6 such trials.

Method

Design and materials. The design included 4 basic triad classifications derived from the 2 × 2 factorial combination of E (good vs. bad) and C (concrete vs. abstract). There were 4 shift conditions: shifts on E only, C only, both dimensions simultaneously, or no shifts on either dimension. Within each of the shift conditions there were 8 groups, with 2 groups representing each of the 4 triad classifications. One of these 2 groups received a shift after a block of 4 prior PI buildup trials, while the other group received a shift after 8 such trials. This design defined the 32 groups that were required to investigate the PI buildup and release phenomena.

In order to obtain the words needed for each of the 4 triad conditions, 2 groups of 30 introductory social psychology *Ss* rated words on each of 2 scales, good–bad and concrete–abstract, since available norms contained few words with ratings on both these dimensions. A total of 160 nouns were selected from available norms of C and E for rating by these *Ss*. Each *S* received a 40-page booklet containing 80 of these critical words (20 from each E × C combination as subjectively determined by *E*) plus 80 filler words. Each of the 2 groups of 30 *Ss* received a different set of 80 critical words but the same 80 filler words. Each page of the booklet contained 4 words with 2 7-point rating scales, concrete–abstract and good–bad, below each word with the former scale always located above the latter for each of the words rated. The booklets were constructed such that 2 critical and 2 filler words appeared in random order on each page. Five random page orders were employed yielding 5 different ordered booklets for each of the 2 sets of words.

Means and standard deviations were calculated from the ratings for each of the critical words and were used to select and place them into one of the following classes: good concrete (GC), bad concrete (BC), good–abstract (GA), and bad abstract (BA). Good was arbitrarily defined as a mean rating ranging 1.00–3.20 on the good–bad scale while bad had ratings ranging 4.80–7.00. Similarly, concrete was defined as mean ratings ranging 1.00–3.20 on the concrete–abstract scale, while abstract words had mean ratings ranging 4.80–7.00. A total of 27 nouns were chosen from each E × C combination. The overall mean ratings of the nouns chosen for each condition were as follows: for GC, 2.09 and 1.63; GA, 2.13 and 5.72; BC, 5.73 and 1.86; and BA, 5.81 and 5.41. The 27 nouns selected for each triad condition were also matched as closely as possible across classes for Thorndike–Lorge (1944) frequency. In constructing word triads, care was taken such that, within a triad there (*a*) was minimum interitem association, (*b*) was variable frequency, (*c*) were different initial letters and variable number of syllables for the words, and (*d*) were variable activity and potency connotative values, as subjectively determined by *E*.

Subjects and procedure. Eight *Ss* were tested in each of 32 groups giving a total of 256 *Ss*. All were undergraduate volunteers enrolled during the summer term at Southern Illinois University at Carbondale. For each of the 32 conditions there

were 2 random presentation orders of the triads since complete counterbalancing would have required too many groups. For each of the conditions, these orders were determined in the following manner. First, 2 random orders for each of the no-shift control groups were defined. The shifted groups then received the same triads in the same order as their preshift controls on preshift trials and the same triads in the same order as their postshift controls on the trial of shift in materials and trials following. Within each condition, 4 *Ss* received each random-order replication. The *Ss* were tested individually and randomly assigned to the 32 groups in the order they came into the lab, with the restriction that the $n + 1$ *S* was not assigned to a group until the nth *S* had been assigned to every other group.

Each word was mounted on a 2 × 2 in. slide and rear-projected by means of a Kodak Carousel projector. After *S* was seated in the experimental room, the typical instructions for the Brown–Peterson (Brown, 1958, Peterson & Peterson, 1959) STM paradigm were read. Two 20-sec. practice periods of backward counting by 3s were then given prior to the start of the experiment to familiarize *S* with the counting task. A trial began with a 2-sec. presentation of an asterisk as a signal that the first word of a triad was about to be presented. Following this signal, the 3 words of the triad were presented individually for 1 sec. each with a projector change time of about .75 sec. The *S* was instructed to read each word aloud as it was presented. Following the last word of the triad, a 3-digit number (from a random-number table) appeared on the screen and remained for a duration of 20 sec. During this time, *S* was required to count backwards aloud from this number by 3s paced by a metronome set at a 1-sec. rate. After 20 sec. of counting, a question mark appeared on the screen and remained for 8 sec. as a signal to *S* to orally recall the words of the previously shown triad. Immediately after the recall interval, an asterisk once again appeared on the screen, signaling the beginning of the next trial. All *Ss* received a total of 9 trials in this manner. The presentation sequence was co[...] by a Gerbrands tape timer.

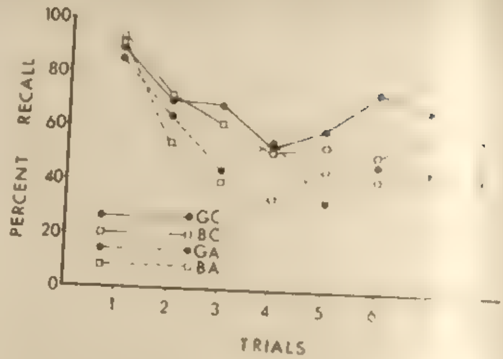


FIGURE 1. Percentage of recall as a f[...] trials for each of the 4 triad conditions. ([...] cluded are those that received no change [...] sion through the first 8 trials, $n = 40$: [...] inhibition buildup. Abbreviations: GC [...] concrete, BC = bad–concrete, GA = good [...] and BA = bad–abstract.)

Scoring. The *E* recorded the recalled [...] as well as intrusions on each trial. In [...] the data, *S* received 1 point for each w[...] rectly recalled regardless of the order [...] resulting in a possible total of 3 points.

TABLE 1

ANALYSIS OF VARIANCE SUMMARY FOR TRIAD CONDITIONS OVER THE FOUR INITIAL TRIALS: PROACTIVE INHIBITION BUILDUP

Source	*df*	*F*
Between *Ss*		
Evaluation (E)	1	5.35
Concreteness (C)	1	22.33*
E × C	1	1.22
Error (*MS*)	252	(.87)
Within *Ss*		
Trials (T)	3	131.40*
E × T	3	3.37
C × T	3	7.15*
E × C × T	3	2.94
Error (*MS*)	756	(.59)

* $p < .01$.

RESULTS

Proactive inhibition buildup. The initial analysis was a 2 × 2 × 4 mixed design, with C (concrete or abstract), E (good or bad), and trials (1–4) as the repeated factor. The results of this analysis are given in Table 1.

In terms of percentage of recall, abstract triads (59.3%) were significantly lower overall than concrete (68.3%), and performance, summed over all conditions, dropped from Trial 1 through Trial 4: 90%, 63%, 55%, and 48%, respectively. The significant C × Trials interaction indicated differential performance across trials for the abstract and concrete triad groups with percentage of recall dropping more rapidly, to a lower final level, for the former. Simple main effect trials analyses for each triad condition indicated significant PI buildup across trials for all 4 triad conditions with the following *F* ratios: GC, $F(3, 756) = 25.54$, GA, $F(3, 756) = 31.53$, BC, $F(3, 756) = 30.63$, and BA, $F(3, 756) = 53.56$, all $ps < .01$. Further analysis

dicated that the groups did not differ .. each other on Trial 1. The percent ill scores for the 4 triad conditions over first 8 trials for those groups not re- ing a shift in triad class on Trial 5 can seen in Figure 1.

)ne additional analysis, a 2^3 between-*S*s .ign, which compared the performance the 4 E × C triad conditions on Trial 4 roups that received 4 PI buildup trials) ith that on Trial 8 (groups that received PI buildup trials) failed to yield any significant effects at the .01 level and thus indicated no further PI buildup with ad- itional trials.

Proactive inhibition release. A 2^5 be- tween-*S*s design was used to examine the release phenomenon. This design included the 2 levels of E and C prior to a shift in triad class (preshift), 2 levels of each dimension following a shift (postshift), and trial of shift (5 or 9). Since no main effect of trial of shift or any interactions with this factor were significant (all *F* ratios < 1), the data were collapsed over this factor and are given in Table 2.

This analysis yielded 4 significant effects: a postshift C main effect, $F(1, 224) = 20.84$, a Preshift × Postshift C interaction, $F(1, 224) = 11.60$, a Preshift × Postshift E interaction, $F(1, 224) = 35.45$, and a triple Preshift × Postshift E × Postshift C interaction, $F(1, 224) = 11.60$, all $ps < .01$, with *MS* error = .74. The 2 double interactions can best be seen by reference to Tables 3 and 4.

TABLE 2

Percentage of Recall for Each Pre- × Postshift Manipulation

Preshift condition	Postshift condition: Good, Concrete	Good, Abstract	Bad, Concrete	Bad, Abstract
Good				
Concrete	58.3	60.3	79.3	87.7
Abstract	75.0	31.3	77.0	62.7
M	66.7	45.8	78.1	75.2
Bad				
Concrete	73.0	66.7	62.7	41.3
Abstract	79.3	75.0	77.0	25.0
M	76.1	70.8	69.8	33.1

TABLE 3

Percentage of Recall for Preshift × Postshift Concreteness: Proactive Inhibition Release

Preshift	Postshift: Concrete	Postshift: Abstract
Concrete	68.3	64.0
Abstract	77.1	48.0
M total	72.7	56.1

From Table 3 it can be seen that recall was substantially better with concrete than abstract postshift triads and that shifted groups exhibited better recall than their appropriate nonshifted controls. Simple main effect analyses comparing the groups shifted with their appropriate nonshifted controls revealed that only the groups shifted from concrete to abstract triads recalled significantly more than their post-shift controls, $F(1, 224) = 9.50$, $p < .01$, while groups shifted from abstract to concrete did not differ from their nonshifted controls. Therefore, significant C shifts were restricted to postshift abstract triads.

Reference to Table 4, the Preshift × Postshift E interaction, also shows superior performance for the shifted E groups compared to appropriate nonshifted controls. However, this interaction also interacted with postshift C as seen by referring to Table 2 where percent recall scores are given collapsed over preshift C. Analysis of the simple Preshift × Postshift E interaction at each level of postshift C indicated a significant interaction effect only when postshift triads were abstract, $F(1, 224) = 43.24$, $p < .01$. It should

TABLE 4

Percentage of Recall for Preshift × Postshift Evaluation: Proactive Inhibition Release

Preshift	Postshift: Good	Postshift: Bad
Good	56.2	76.7
Bad	73.2	51.5
M total	64.7	74.1

perhaps be noted that the direction of the interaction for the postshift concrete triads, although not significant, also indicated better recall for the groups shifted on E than those not so shifted. Furthermore, inspection of the Preshift × Postshift E interactions indicated similar differences between shifted groups and their respective postshift controls, i.e., relatively symmetrical shift effects regardless of direction (from good to bad or vice versa). Consonant with these observations, simple main effect analyses comparing groups shifted on E with appropriate postshift control groups indicated significant shift effects from bad to good and good to bad with postshift abstract triads, $F(1, 224) = 12.16$, and $F(1, 224) = 33.74$, respectively, both $ps < .01$, but no significant effects in either direction with postshift concrete triads.

Another way to evaluate the triad shift data was to partition the groups into 4 different shift conditions (no shift, shift on E only, C only, and both E and C) and analyze recall on the 4 E × C postshift triad types with the additional factor of trial of shift (5 or 9). In this analysis, the effects of shifts on a particular dimension were isolated from shifts on other dimensions, whereas in the previous analysis, the effects of shifts on one dimension were averaged over shifts and no shifts on the other dimension.

Neither postshift E, trial of shift, nor any interactions involving these factors were significant. Significant effects were obtained for shift condition, $F(3, 224) = 16.63$, postshift C, $F(1, 224) = 20.84$,

TABLE 5

PERCENTAGE OF RECALL FOR THE VARIOUS SHIFT CONDITIONS UNDER BOTH LEVELS OF POSTSHIFT CONCRETENESS

Shift condition	Postshift concreteness	
	Concrete	Abstract
No shift	60.0	28.0
Concreteness (C) only	76.0	51.0
Evaluation (E) only	76.0	68.7
Both C and E	79.3	77.0

Note. Data averaged over Trials 5 and 9.

and the Shift × Postshift C interaction, $F(3, 224) = 4.16$, all $ps < .01$, with MS error $= .74$. Mean percentage of recall for this analysis, collapsed over postshift E and trial of shift, is given in Table 5.

Simple shift conditions main effect analyses under each level of postshift C indicated a significant effect only under postshift abstract, $F(3, 224) = 18.27$, $p < .01$, consistent with the previous analyses. Pair-wise comparisons of the various shift conditions under postshift abstract yielded significant differences between each of the shifted conditions and the no shift condition as well as a significant difference between the double (E-C) shift condition and the C only condition with an overall .05 alpha level.

Discussion

With regard to the question of differential PI buildup, the data clearly show differences in such buildup as a function of C but not E. Previous studies have tended to suggest no differential buildup for words on either end of the evaluative dimension (Turvey, 1968; Wickens & Clark, 1968) and this is confirmed by the results of the present study. The finding of greater PI accumulation over trials for abstract triads compared to concrete triads is likewise consistent with available evidence (Paivio & Begg, 1971; Wickens & Engle, 1970). In general, a facilitative effect of concreteness and or imagery has been demonstrated in a variety of verbal learning situations where this dimension has been manipulated (Paivio, 1971) and thus, would not be unexpected here.

If PI release is defined in terms of the difference in recall performance between a group shifted on one or more dimensions and its postshift control, then reliable release was demonstrated for shifts on both E and C as indicated by their respective Preshift × Postshift interactions.

Reliable release for shifts on E simply replicates the results of previous studies (Turvey, 1968; Wickens & Clark, 1968) and was not unexpected. One finding of interest in this regard however, was that shifts on E interacted with postshift C such that the difference between shifted and appropriate nonshifted controls was significant only under the postshift abstract condition. If, as Wickens and Engle (1970) implied, abstract

and concrete words are not encoded differently and represent a dimension of relative unimportance in an STM experiment, then release should have obtained with shifts on E irrespective of the level of C. If instead, imaginal encoding offsets the interference effects of a verbal code (such as connotative evaluation), as has been suggested by Paivio and Begg (1971), then little release would be expected with shifts on E when the words are highly concrete contrasted with a large release with shifts on E when the words are highly abstract. The results of the present study are entirely consistent with this latter prediction.

Furthermore, the present study also indicated reliable release for shifts on C (for concrete to abstract triad shifts) which Wickens and Engle (1970) failed to obtain. Several reasons for such a discrepancy might be given. The present study contained a number of procedural differences from that of Wickens and Engle such as the amount of time given for recall (6 vs. 8 sec.), requesting *S*s to recall triad items in serial order of presentation vs. free-recall instructions, and the presentation of all 3 items of the triad at one time for 3 sec. vs. presentation of each item of the triad separately for 1 sec. each. One might argue that this last difference is crucial in the sense that more total time was available in the present experiment than in the Wickens and Engle study, to encode the words of a triad prior to the distractor task (due to projector change time), thus allowing more time for image formation. However, as Paivio (1971) has pointed out, evidence would seem to suggest that image latency to single words is in the order of .6 sec. which would tend to minimize the importance of such a distinction.

A major difference between the studies was the control for E. This may be important in the sense that a reliable dimension (E) was available in the present experiment for encoding regardless of the C manipulation, whereas in the Wickens and Engle (1970) experiment, no consistent dimension other than C was available.

When one considers the very nature of the C dimension, it becomes apparent that the situation with respect to shifts on this dimension is considerably more complex than shifts on other dimensions. Concrete and abstract words may be considered alike in that both may be encoded verbally and, as such, there should be no differential encoding. However, concrete words provide for an additional encoding mechanism, imagery. This dual encoding possibility for concrete words does not exist for other encoding dimensions investigated by Wickens (1970) and his colleagues and makes this dimension unique in that respect. Thus, when one shifts from abstract to concrete words, an additional encoding mechanism is made available to the one employed preshift. Performance on such a shift might reasonably be expected to improve relative to that preshift but not necessarily be significantly different from a control which always had the opportunity to employ a dual encoding mechanism. A shift from concrete to abstract (dual encoding to single encoding mechanism) should result in a decline in performance relative to that preshift due to the loss of an encoding mechanism. Performance on such a shift might not be expected to differ from an abstract control unless one of the following prevailed. First, if substantial interference buildup occurred for images in addition to such buildup for a verbal code, then a shift from concrete to abstract items would eliminate imaginal interference resulting in PI release relative to an abstract control. Such a hypothesis seems unlikely since it would also predict greater PI buildup over trials for concrete vs. abstract triads, a situation which has been repeatedly shown to be the reverse.

The second possibility, which would predict a difference in performance in favor of a group shifted from concrete to abstract relative to an abstract control, as well as superior performance preshift for concrete items, involves Paivio and Begg's (1971) offsetting interference hypothesis. If the availability of an imaginal code preshift offsets the interference effects generated by a verbal code, then a shift from a dual encoding mechanism to a single encoding mechanism (verbal) should result in a performance decrement from that demonstrated preshift but to a level somewhat above an abstract control where interference buildup along a verbal code would have been great. This latter possibility would predict an empirical demonstration of PI release. The C shift results of the present study would appear consistent with this offsetting interference hypothesis and Paivio's (1971) dual encoding theory for concrete words.

REFERENCES

Brown, J. A. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, 10, 12–21.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *Measurement of meaning.* Urbana: University of Illinois Press, 1957.

PAIVIO, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.

PAIVIO, A., & BEGG, I. Imagery and associative overlap in STM. *Journal of Experimental Psychology,* 1971, **89**, 40–45.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement,* 1968, **76** (1, Pt. 2).

PETERSON, L. P., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology,* 1959, **58**, 193–198.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

TURVEY, M. T. Analysis of augmented rec short-term memory following a shift in con tion. *British Journal of Psychology,* 1968, 131–137.

WICKENS, D. D. Encoding categories of wo An empirical approach to meaning. *Psycholog Review,* 1970, **77**, 1–15.

WICKENS, D. D., BORN, D. G., & ALLEN, C Proactive inhibition and item similarity in sho term memory. *Journal of Verbal Learning an Verbal Behavior,* 1963, **2**, 440–445.

WICKENS, D. D., & CLARK, S. Osgood dimensi as an encoding class in short-term mem *Journal of Experimental Psychology,* 1968, 580–584.

WICKENS, D. D., & ENGLE, R. W. Imagery and abstractness in short-term memory. *Journal of Experimental Psychology,* 1970, **84**, 268–27

(Received June 29, 1973)

(*Continued from page 313*)

Hypothesis Behavior as a Function of Amount of Pretraining. Raymond M. White, Jr. (Department of Psychology, Miami University, Oxford, Ohio 45056).

The Locus of the Retention Differences Associated with Degree of Hierarchical Conceptual Structure Benton J. Underwood (Department of Psychology, Northwestern University, Evanston, Illinois 60201), John J. Shaughnessy, and Joel Zimmerman.

The Effect of Number of Stimuli on the Steady-State Generalization Gradient: A Test of the Summation Hypothesis. John C Malone, Jr. (Department of Psychology, University of Tennessee, Knoxville, Tennessee 37916).

Effects of Evaluation Apprehension on Memory over Intervals of Varying Length. Russell G. Geen (Department of Psychology, 209 McAlester Hall, University of Missouri-Columbia, Columbia, Missouri 65201).

Strength of Auditory S–R Compatibility as a Function of Task Complexity. James R. Callan (Department of Psychiatry and Behavioral Sciences, University of Oklahoma Health Sciences Center, P.O. Box 26901, Oklahoma City, Oklahoma 73190), Diane Klisz, and Oscar A. Parsons.

Target Location Aftereffects for Various Age Groups. Benjamin Wallace (Department of Psychology, Western Illinois University, Macomb, Illinois 61455) and Scott P. Anstadt.

Extended Mediation in Children's Paired-Associate Learning. Lewis P. Lipsitt (Walter S. Hunter Laboratory of Psychology, Brown University, Providence, Rhode Island 02912) and Judith E. Sims-Knight

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 321-323

SHORT REPORTS

PRONUNCIATION AND APPARENT FREQUENCY IN A BETWEEN-SUBJECTS DESIGN [1]

LARRY WILDER,[2] JOEL R. LEVIN, ELIZABETH S. GHATALA, AND SANDRA McNABB

University of Wisconsin—Madison

College students were administered a list of middle frequency words, in which individual words were presented 1–6 times. Half of the *S*s pronounced the list, while the other half remained silent. On a subsequent frequency judgment task, pronunciation *S*s failed to differ significantly from silent *S*s on mean judgments of items presented only once; however, there were differences on these items as indexed by variability and mean total correct measures. Analysis of judgments of *zero* items (not seen on the study list) indicated differences between silent and pronunciation *S*s on all 3 measures, while the pronunciation effects tended to be eliminated on the items presented more than once. These results were contrasted with previous findings and explained within a frequency theory perspective.

Situational frequency, as described by Ekstrand, Wallace, and Underwood (1966), has been shown to be a principal attribute in verbal discrimination learning (VDL) and recognition memory. In VDL, situational frequency is in part a function of rehearsal of the correct response, which leads to an increased frequency differential between the correct and incorrect members in a pair. In recognition memory, frequency of exposure during the learning trial is a significant factor affecting performance (Underwood & Freund, 1970).

Spoken rehearsal has been found to be superior to control (i.e., silent) performance in VDL (Wilder, 1971). Frequency theory can account for this finding if it is assumed that pronouncing during rehearsal increases the situational frequency differential between the correct and incorrect item. Hopkins, Boylan, and Lincoln (1972) tested this assumption in a frequency judgment task. In one experiment, they found no difference in mean frequency judgments between *S*s who pronounced all words on the study trial and *S*s who did not pronounce (Experiment 3). However, when this manipulation was made within *S*s, pronunciation did increase apparent frequency (Experiment 4). That is, when *S*s pronounced some words but not others in the same list, they judged the pronounced words to have a higher frequency of occurrence. These results led the authors to conclude that pronunciation influenced (increased) apparent frequency only "if *S* has also had experience with silently studied items in the same experimental context [Hopkins et al., 1972, p. 112]," and at the same time permitted a frequency explanation of the facilitative effect of pronouncing the correct response in VDL.

However, Underwood (personal communication, November 1971) has justly pointed out that an *absolute* mean increase in the situational frequency associated with the "correct" response in VDL, or with previously seen items in recognition memory, is not the only way in which an increased frequency differential can occur. For example, if pronunciation *S*s exhibit lower variability than silent *S*s on list items with the *same* situational frequency, then an equivalent increased frequency differential among items with *different* situational frequencies would be the result. Thus, while Hopkins et al. (1972) found no mean differences in an independent groups design, it is possible that pronunciation increases item discriminability by reducing the variance of mean frequency judgments for items of the same situational frequency. In other words, while apparent frequency units may not be added as a function of pronouncing the items on a study list, more consistent frequency estimates of items with the same situational frequency would be expected to produce the same net result in VDL and recognition memory tasks.

Accordingly, the present experiment was intended to assess the effects of pronunciation on frequency judgment variability in a between-*S*s design. In addition, a measure found to be correlated with, but more sensitive than, within-*S* variability—total number of correct frequency judgments (Ghatala & Levin, in press)—was considered. The experiment itself was similar to Hopkins et al. (1972), Experiment 3; however, in addition to just study list items, a number of filler items were presented during the second (test) list. It was predicted that if pronunciation decreases variability on items previously seen, this would result in increased discriminability between these items and items not presented, as in recognition memory.

---

[1] This research has been published by the Wisconsin Research and Development Center for Cognitive Learning, which is supported in part by the U.S. Office of Education, Department of Health, Education, and Welfare, Center No. C-03, Contract OE 5-10-154. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred. We are indebted to B. J. Underwood for his helpful comments throughout this study. The assistance of Janis Van Brocklin and Billie Albrecht in preparing the article is also appreciated.

[2] Requests for reprints should be sent to Larry Wilder, Wisconsin Research and Development Center for Cognitive Learning, 1025 W. Johnson, Madison, Wisconsin 53706.

Method. A total of 42 upper division University of Wisconsin communication arts students participated in the experiment for partial course credit.

The stimuli consisted of 2 lists of middle-frequency (10–20 on the general Thorndike–Lorge, 1944, list) 2-syllable nouns. The study list contained 90 words. Fifty of the words were presented once; 25 were presented twice; 10 were presented 3 times; 3 were presented 4 times; one was presented 5 times; and one was presented 6 times. In all, there were 153 presentations. The list was divided so that each word occurred equally often in equal divisions of the list (e.g., for the one word that occurred 6 times, the list was divided into sixths). The second list was the test list, which included all of the 90 words from the study list plus 10 additional words that had not been presented on the study list.

The first list was presented on a memory drum at a 2-sec. rate. Two groups of *S*s were used. A pronouncing group was told:

> This is an experiment on word memory. You will be shown a long list of words. Some words will occur once, while others will occur twice on up to six times. Your task is to look at each word carefully and pronounce it. The list is too long to actually count and remember all of the words, so just try to get an impression of the number of times each word has occurred.

A silent group was given the same instructions, but they were told only to look at each word carefully. A test list was presented once on a memory drum at a 3.5-sec. rate to both groups. All of the *S*s reported verbally how many times they had seen the word before. They were told that they could report *zero* for any item that they had not seen previously.

Results. A summary of *S*s' frequency judgment performance on the *one* and *zero* items, as well as the average number of correct frequency judgments (out of 100), is presented in Table 1 with the corresponding statistics based on pronunciation–control differences. Since the predictions derived from frequency theory were all quite explicit (i.e., pronunciation should result in larger means and smaller variability on *one* items, in both smaller means and variability on *zero* items, and in a greater number of correct frequency judgments), all comparisons made were directional and thus the significance probabilities reported in Table 1 are 1-tailed.

TABLE 1

SUMMARY MEASURES AND ONE-TAILED SIGNIFICANCE PROBABILITIES CORRESPONDING TO PRONUNCIATION–CONTROL DIFFERENCES

Measure	Control	Pronunciation	t (40)	p
		One items		
M	1.090	.947	−1.46	—[a]
Variance	.603	.456	−1.84	<.05
		Zero items		
M	.348	.115	−3.63	<.001
Variance	.326	.152	−2.37	<.025
		All items		
Total correct	45.19	55.10	3.18	<.005

[a] Wrong direction.

Mean apparent frequencies were computed for each *S* by averaging his frequency judgments on the 50 *one* items. The same was done for each *S* on the 10 *zero* items. The mean in Table 1, therefore, represents the average mean across *S*s in the same condition. As may be seen, the difference between the pronunciation and control conditions for mean judgments of *one* items was in the wrong direction and consequently nonsignificant, which corresponds to the Hopkins et al. (1972) result. Mean judgments for *zero* items did result in significant differences between the 2 conditions, however.

Within-*S* variances were obtained on *one* and *zero* items by computing each *S*'s mean squared deviation (about his apparent frequency mean). The entries in Table 1 represent the average of these within-*S* variances within each condition. Pronunciation resulted in lower average within-*S* variances, particularly on the *zero* items.

The total correct measure was computed by counting the total number of times *S*'s frequency judgment corresponded exactly to the appropriate actual frequency. It can be seen that pronunciation *S*s averaged about 10 more correct responses than did control *S*s (this effect was found to hold for both *one* and *zero* items when analyzed separately).

Analysis of means and variances of items presented more than once detected only one other statistically significant pronunciation–control difference: lower within-*S* variances for pronunciation *S*s on the *four* items, t (40) = 2.12, p < .05.

Discussion. These results are consistent with those of Hopkins et al. (1972) in that there was no significant difference between pronunciation and silent *S*s on mean frequency judgments. However, going beyond mean frequency judgments, significant differences between silent and spoken performance were detected in a between-*S*s design. There was less variability within *S*s, as well as a greater mean total correct, for *S*s who pronounced the *one* items. Further, the prediction that *S*s who pronounced the items on the study list were better able to discriminate items they had not seen before was confirmed.

From these results it would appear that Hopkins et al. (1972) were premature in discounting the effects of pronunciation in a between-*S*s design, and that frequency theory can account for these results if it is assumed that reduced variability contributes to improved frequency discriminations. Also, it is not surprising that these effects tend to be eliminated when items are presented more than once, since Ekstrand et al. (1966, p. 575) have suggested, according to Weber's law, that adding frequency

its to items with greater situational frequency should be less noticeable.

REFERENCES

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, 73, 556–578.

Ghatala, E. S., & Levin, J. R. Developmental differences in frequency judgments of words and pictures. *Journal of Experimental Child Psychology*, in press.

Hopkins, R. H., Boylan, R. J., & Lincoln, G. L. Pronunciation and apparent frequency. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 105–113.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

Underwood, B. J., & Freund, J. S. Testing effects in recognition of words. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 117–125.

Wilder, L. Spoken rehearsal and verbal discrimination learning. *Speech Monographs*, 1971, 38, 171–176.

(Received October 28, 1972)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 323–326

COGNITIVE PROCESSING OF LINEAR ORDERINGS[1]

KARL W. SCHOLZ[2]
Indiana University

GEORGE R. POTTS
Dartmouth College

This study attempted to discriminate between 2 theories about strategies Ss use when trying to remember a linear ordering. Rating-scale theory argues that *Ss* represent the items to be learned as points along an imaginary spatial continuum. End-term anchoring theory suggests that serial orderings are learned from the end points inward. A scaling analysis of proportions correct suggested that neither theory alone is sufficient. The Ss apparently represented ordering information along 2 dimensions, the first representing the distance between items on the imaginary continuum and the second representing the distance of the items from the center of the ordering.

In an attempt to clarify the encoding strategies *Ss* use when trying to remember a linear ordering, Potts (1972) had Ss learn a paragraph that presented an ordering of 4 terms along a meaningful dimension. Such a 4-term ordering (A > B > C > D) can be broken into 6 pairs of terms, 3 of which (the "adjacent" pairs, A > B, B > C, and C > D) represent relations between adjacent elements in the ordering and are therefore essential to the establishment of that ordering. The remaining 3 pairs (the "remote" pairs, A > C, B > D, and A > D) are redundant in the sense that they can be deduced from some subset of the adjacent pairs. Potts found that responses to the remote pairs were both faster and more accurate than responses to the adjacent pairs. This was true even when the remote pairs were never presented and thus had to be deduced from the adjacent pairs. These data were shown to be inconsistent with several possible theories regarding the form in which such information is stored. Two theories were proposed that could account for this result.

The first of these theories, an extension of Huttenlocher's (1968) spatial-imagery model, argues that Ss represent the items to be learned as points along an imaginary spatial continuum and will be referred to herein as *rating-scale theory*. In such a model, forgetting is represented by variability in the placement of the items on this scale. The Ss incorrectly identify a particular pair whenever the variability between the 2 terms comprising that pair is sufficiently large to permit the 2 terms to cross each other on the scale. This model would predict proportion correct on a pair to be a monotonic increasing function of the distance between the 2 terms comprising that pair. Since the terms of a remote pair would have to be placed farther apart on the scale than the terms of any adjacent pair necessary to deduce it, the theory would necessarily predict better performance on the remote pairs than on the adjacent pairs.

An alternative explanation of Potts' (1972) result is provided by research (DeSoto & Bosley, 1962; Feigenbaum & Simon, 1962; Wishner, Shipley, & Hurvich, 1957) suggesting that serial lists are learned from the end points inward, with the first and last terms serving as "anchors" for the other terms in the list. If one accepts the argument that there is less variability in the identification of the ordinality of end terms, it follows that performance on the remote pairs should surpass performance on the adjacent pairs since, on the average, the remote pairs contained more end terms than did the adjacent

[1] This work was supported by U. S. Public Health Service Grant MH-16817. The authors wish to thank Henry Halff, Larry Jones, and Janet Walker for their critical appraisal of the manuscript.

[2] Requests for reprints should be sent to Karl W. Scholz, who is now at the Department of Psychology, University of Illinois, 602 South Sixth Street, Champaign, Illinois 61820.

pairs. In what follows, this will be referred to as an *end-term anchoring theory*.

A straightforward test of these 2 alternatives would involve examining proportion correct as a function of distance, with number of end terms held constant. If proportion correct still increased as a function of distance under this condition, we would conclude that end-term anchoring was not important. If, on the other hand, holding the number of end terms constant eliminated the effect, the importance of end terms would be affirmed. Unfortunately, with the 4-term orderings employed by Potts (1972), the effects of distance and number of end terms were completely confounded. The present experiment examined proportions correct on 6-term orderings in an attempt to unconfound these 2 factors.

Method

Subjects. The *S*s were 40 Indiana University undergraduates who participated to fulfill a course requirement. Each *S* served in a single 30-min. session.

Apparatus. The experiment was conducted in the Mathematical Psychology Laboratory at Indiana University. Stimuli were presented on an 11-in. Sony video monitor that was controlled by an Applied Digital Data Systems MRD-280 video character generator. True–false responses were obtained using a 2-button response box located immediately below each video monitor. Each button was appropriately labeled with the word *true* or *false*. Stimulus presentation and response collection were controlled by an IBM 1800 process-control computer.

Materials. Stimuli consisted of 9 English paragraphs presented one at a time on the video monitor. Each paragraph represented a linear ordering among 6 items along a unique dimension. The syntax for all 9 paragraphs was identical and presented the intended ordering as 5 pair-wise relations. Several researchers (Clark, 1969; Huttenlocher, 1968; Jones, 1970; G. R. Potts and K. W. Scholz, manuscript in preparation) have shown that differences in the linguistic markedness of certain oppositional adjective pairs affect the speed with which sentences using these adjectives are processed. In order to avoid any unintentional confounding by lexical marking, the 5 relational terms used in each paragraph were identical. In order to reduce variability in the case of paragraph comprehension, the surface structure of each paragraph was held constant. Each of the 9 paragraphs had the structure, "Mary is smarter than Jim, and Jim is smarter than Betty. Betty is smarter than Tom. Tom is smarter than Alice, and Alice is smarter than Bob." If the capital letters A–F are used to represent the 6 items, and the symbol > used to represent the relation, then the surface structure of each paragraph can be represented as "A > B and B > C. C > D. D > E and E > F."

The 6-term orderings used in the present experiment permit the testing of 15 distinct pairs of terms. Each test statement consisted of a single sentence such as "Jim is smarter than Alice." For each true test sentence of the form "A is smarter than B," there was a corresponding false sentence of the form "B is smarter than A" that presented the same 2 terms in reverse order. In this way, a total of 30 statement types was obtained. Each statement type was tested 3 times in a session. The 90 test statements were randomly assigned to the 9 paragraphs with the constraint that 5 true and 5 false statements followed each paragraph.

Procedure. From 1 to 4 *S*s participated in each experimental session. Prior to the beginning of the session, *S*s were given a brief training trial to acquaint them with the apparatus and procedure. This training trial consisted of presenting *S*s a sample paragraph followed by 2 sample test questions which they were required to answer

Following this training trial, the first of the 9 test paragraphs was presented on the viewing screen. The paragraph remained in view for 90 sec or until all *S*s indicated (by a button press) that they could remember the information in the paragraph. The time required to study each paragraph was not specifically recorded, but spot checks revealed that this time rarely exceeded 60 sec. As soon as all *S*s in a session indicated that they were ready, the screen was erased and the first test statement presented. The *S*s were instructed to respond to the statement (by a button press) as quickly as possible. As soon as all *S*s had responded, the test statement was erased, and the next statement was presented after a 1-sec. intertrial interval. The procedure was identical for the 10 test statements following each of the 9 paragraphs.

Results and Discussion

Mean proportions correct for each of the test pairs, averaged over true and false sentences, are presented in Table 1. The entry in each cell represents the proportion correct on the pair made up of those terms indicated by the row and column headings. Each pair can be characterized according to the number of end terms it has (0, 1, [illegible] its step size (1, 2, 3, 4, or 5), where the [illegible] of a pair refers to the number of adja[illegible] necessary to deduce it. Clearly, step size is a monotonic function of the distance separating the terms of the pairs.

A 2-way analysis of variance (step size by true/false) indicated a significant main effect for step size, F (4, 156) = 9.84, $p < .001$, but not for true vs. false, F (1, 39) = 2.98, $p = .09$. Orthogonal polynomial trend analysis revealed a signifi[illegible] linear trend, F (1, 156) = 29.71, $p < .001$, for step size. Higher order trends were not significant. The observation that performance is better for more widely separated items is consistent with Potts' (1972) data and conforms to the predictions of the rating-scale model. However, the average number of end terms included in the test pairs increases monotonically as a function of step size. Thus, the apparent step size effect could also be explained by

the assumption that end terms are more readily accessible.

Similarly, average proportion correct is higher for pairs that include an end term than for pairs that do not, t (39) = 5.04, $p < .001$. Though this effect is consistent with the predictions of the end-term anchoring theory, it is confounded with step size.

In an attempt to distinguish between the 2 models, proportion correct as a function of step size was examined, with number of end terms held constant. First, proportion correct as a function of step size was examined for those test pairs which did not include any end term (i.e., for the subordering BCDE). The mean proportions correct for pairs with Step Sizes 1, 2, and 3 were .713, .758, and .771, respectively. This represents a significant linear trend, F (1, 78) = 3.84, $p = .05$, which accounted for almost all (89%) of the variance between the cell means of the step size variable. Thus, as predicted by the rating-scale theory, even with the number of end terms held constant at zero, proportion correct is still a monotonic increasing function of step size.

Second, proportion correct as a function of step size was examined for those test pairs that included exactly one end term. The mean proportions correct for pairs with Step Sizes 1, 2, 3, and 4 were .771, .850, .858, and .804, respectively. These proportions correct are clearly not a monotonic increasing function of step size. The quadratic trend in this case was highly significant, F (1, 117) = 18.05, $p < .001$, and accounted for almost all (88%) of the variance between the cell means of the step size variable. The linear trend was not significant, F (1, 117) = 2.33, $p = .13$, and accounted for only 11% of the variance.

The extent to which the present data are inconsistent with the predictions of a rating-scale theory can also be seen by examining the exact predictions of such a theory. Let the first through sixth terms in the ordering be represented by the numbers 1 through 6, respectively; let $D(ij)$ represent the spatial distance between the items i and j on the scale; and let $P(ij)$ represent the proportion correct on the pair consisting of the items i and j. The basic prediction of a rating-scale theory is that $P(ij)$ is a monotonic increasing function of $D(ij)$. In other words, for any i, j, m, and n, $P(ij) = P(mn)$ if and only if $D(ij) = D(mn)$, and $P(ij) > P(mn)$ if and only if $D(ij) > D(mn)$.

If one examines the nature of a linear scale, it becomes apparent that $D(ij) > D(mn)$ whenever $i \leq m$ and $j \geq n$. (The only exception is the trivial case where both $i = m$ and $j = n$, and this case can be ignored in the present analysis.) Thus, a rating-scale theory would predict that $P(ij) > P(mn)$ whenever $i \leq m$ and $j \geq n$. Examination of Table 1 indicates that this prediction is completely consistent with the data for those pairs that do not include either end term (i.e., pairs included in the subordering BCDE), but fails among those pairs that do include one of the end terms. Though performance is best on the pair A > F, which is

TABLE 1

Proportions Correct on 15 Test Pairs

Pair	A	B	C	D	E	F
A	—					
B	.771	—				
C	.858	.738	—			
D	.842	.763	.713	—		
E	.783	.771	.754	.688	—	
F	.871	.825	.875	.842	.771	—

the most distant, performance is better on the pairs A > C and A > D than on the pair A > E; and performance is better on the pair A > C than on the pair A > D. Similarly, performance is better on the pairs D > F and C > F than on the pair B > F.

Thus, it appears that although a rating-scale theory can account for the proportions correct on pairs relating to the inner terms (BCDE), its predictions are incorrect with respect to the proportions correct on pairs which contain one of the end terms. In order to clarify the nature of the discrepancies, a multidimensional scaling analysis was performed on the proportions correct. In doing this, the scaling assumption that proportion correct is a monotonic increasing function of distance was accepted, but the dimensionality of the resulting cognitive space was viewed as an empirical question. The scaling solutions that follow were obtained using TORSCA, a FORTRAN IV program, described by Young and Torgerson (1967).

The rating-scale theory predicts that the cognitive space is unidimensional, since it argues that *Ss* place the terms along a line. Thus, the unidimensional scaling solution based on proportions correct should reflect the true spatial distances between the items, and thus would have to result in those items being ordered correctly on the scale. To the extent that the unidimensional solution does not correspond well to the actual ordering of the terms, it would indicate that the unidimensional rating scale is not an accurate representation of the way *Ss* store or access the ordering.

The unidimensional solution is shown in the top frame of Figure 1. It is clear that this solution does not accurately represent the true ordering of the terms. The stress values for the 1-, 2-, and 3-dimensional solutions were .274, .027, and .001, respectively. The poorness of the unidimensional fit relative to the 2-dimensional fit is readily apparent. In order to investigate the possibility that the 1-dimensional solution shown in Figure 1 was the consequence of entrapment in a local minimum, stress was estimated for the ABCDEF ordering, assuming both equal spacing and spacing that emphasized the remoteness of the end terms. Using the numerical method described by Kruskal

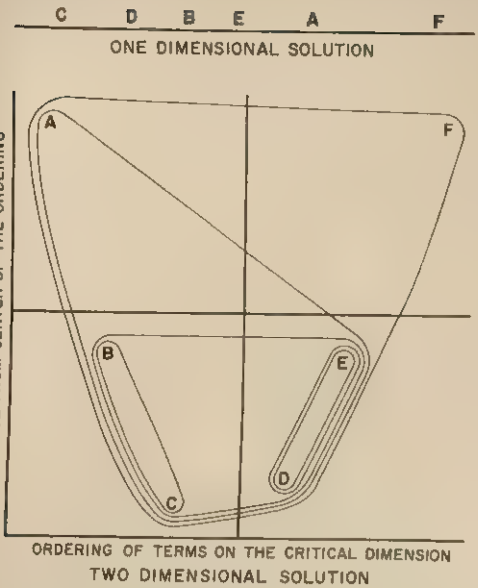


FIGURE 1. The 1-dimensional (upper frame) and 2-dimensional (lower frame) scaling solutions for proportions correct.

(1964), we discovered no set of distances that would yield a stress of less than .2.

The problem with the unidimensional solution becomes apparent when one examines the best 2-dimensional solution, shown in the bottom frame of Figure 1. A pair of easily interpretable orthogonal axes is included in the figure. The projection of points onto one of these axes gives the actual linear ordering of the terms on the critical dimension; the projection on the other axis gives the distance of each item from the center of the ordering. It is clear that the end terms are very different from the inner terms (BDCE) of the ordering on both dimensions. On one dimension, however, these end terms are the items which are farthest apart; on the other dimension, they are closer to each other than they are to any other item.

The clustering analysis presented along with the 2-dimensional solution was obtained using the diameter method. The result again demonstrates the uniqueness of the end terms relative to the other terms in the ordering. Use of the connectedness method yielded minor differences in the clustering among the inner terms, but showed essentially the same effect.

Our interpretation of the 2-dimensional scaling solution is that one dimension represents the distance between items on an imaginary scale. The ordering of the items along this dimension is consistent with a rating-scale theory. The other dimension represents the distance of the items from the center of the ordering. On this dimension, items more distant from the center of the ordering (hence nearest the ends) are grouped together. The grouping is consistent with the assumption of end-term anchoring theory that end terms are learned best.

The conclusion drawn from the results is that neither a rating-scale theory nor an end-term anchoring theory is adequate in itself as an explanation of how *S*s process linear orderings. The predictions each theory would make about performance were described earlier. The results suggest that the theories be integrated. Apparently, the accuracy of *S*'s attempt to identify the truth or falsity of a test pair is a function of the ease with which the relative position of each item can be identified (as in the rating-scale theory). In addition, position identification is facilitated if the item is an end term in the complete ordering (as in an end-term anchoring theory). These 2 components could reflect either differences in how the information is structured in memory, or differences in the strategies by which the information is retrieved. Further work would be necessary to discriminate these 2 alternatives.[3]

[3] We note that there is a superficial similarity between our task and various psychophysical scaling tasks. Our task was designed to investigate memorial representation and retrieval of linearly ordered meaningful material, and as such we feel that an attempt to relate our findings to existing psychophysical models would be unwarranted and potentially misleading.

REFERENCES

CLARK, H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **76**, 387–404.

DESOTO, C. B., & BOSLEY, J. J. The cognitive structure of a social structure. *Journal of Abnormal and Social Psychology*, 1962, **64**, 303–307.

FEIGENBAUM, E. A., & SIMON H. A. A theory of the serial position effect. *British Journal of Psychology*, 1962, **53**, 307–320.

HUTTENLOCHER, J. Constructing spatial images: A strategy in reasoning. *Psychological Review*, 1968, **75**, 550–560.

JONES, S. Visual and verbal processes in problem-solving. *Cognitive Psychology*, 1970, **1**, 201–214.

KRUSKAL, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964, **29**, 115–129.

POTTS, G. R. Information processing strategies used in the encoding of linear orderings. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 727–740.

WISHNER, J., SHIPLEY, T. E., & HURVICH, M. S. The serial-position curve as a function of organization. *American Journal of Psychology*, 1957, **70**, 258–262.

YOUNG, F. W., & TORGERSON, W. S. TORSCA, a FORTRAN IV program for Shepard & Kruskal multidimensional scaling analysis. *Behavioral Science*, 1967, **12**, 498.

(Received January 18, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 327–329

B–A LEARNING AS A FUNCTION OF DEGREE OF A–B LEARNING

C. MICHAEL LEVY[1] AND DOROTHY D. NEVILL

University of Florida

Two studies were conducted to examine the backward-learning function. In Experiment I, 5 groups of 20 *S*s each learned an 8-item A–B list where the A and B terms were single alphabetical or numerical symbols. Learning was to a criterion of either 0, 2, 4, 6, or 8 correct anticipations. Subsequently, *S*s were trained to an 8/8 criterion on the B–A list. Experiment II was designed to control for the required temporal delay between A–B and B–A learning in Experiment I. The data revealed a number of similarities between the A–B and B–A functions, but provided no support for the associative symmetry hypothesis or for one of its important variants.

Systematic studies of the acquisition of backward associations are rare. Jantz and Underwood's (1958) study today has only heuristic value due to methodological difficulties resulting from their use of nonsense syllable–adjective pairs. In general, studies using these materials have not addressed adequately the problem of stimulus–response availability. It could be argued that on the completion of forward (A–B) learning an *S* would have learned B and the association between the A and B terms. But since the A term had not been specifically learned as such, and indeed *S* may not have attended to all of the nominal attributes of the A term, it would not be as available as the previously learned B. Several methods have been used to control for availability of stimulus–response terms, including prelearning and use of double-function lists. However, the former method has the possibility of producing interitem connections which could greatly modify subsequent paired-associate learning and the latter results in lower rates of B–A learning than when other types of lists are learned. Consequently, a simpler solution to the problem of the unequal availability of A and B terms is to avoid using highly complex stimuli (e.g., nonsense syllables) and to use instead very simple, highly available items such as single digits or letters of the alphabet (Ekstrand, 1966).

A recent study (Baumeister & Campbell, 1971) attempted to avoid the design problems discussed above by using black and white drawings of highly familiar objects in order to equate their availability as both stimulus and response terms. Baumeister and Campbell observed an orderly relationship between the forward and backward associative functions, although the relationships were not identical for the normal and retarded children studied. In their first experiment, where the focus was the amount of B–A recall as a function of the degree of A–B learning, the number of A–B trials was the principal variable manipulated. In contrast, the proportion of the A–B list correctly anticipated in the present studies was varied in order to examine its influence upon the course of B–A learning.

EXPERIMENT I

Method. The list were composed of 8 single digit and letter pairs selected to maximize item availability. The digits 2–9 were chosen, all of which rank in the upper quartile in association value for the numbers 0–100. The letters used were B, M, S, T, H, D, R, and F, which are among the 9 letters highest in meaningfulness. The materials were presented on a memory drum at a 1:1 rate to minimize intratrial rehearsal and to prevent *S*s from "running through" the A–B sequence during B–A learning (Ekstrand, 1966).

During Stage 1, *S*s engaged in forward anticipation learning. Group 8 learned a list to a criterion of 8 out of 8 correct responses. Subjects in Groups 6, 4, and 2 were taken to criteria of, respectively, 6 out of 8, 4 out of 8, or 2 out of 8 correct responses. Group 0 was not presented with the A–B list, but proceeded directly to Stage 2.

Stage 2 consisted of backward anticipation learning. A 1-min. period intervened between Stages 1 and 2 during which the instructions were given for the second stage. All groups learned the B–A list to a criterion of 1 perfect errorless anticipation.

List order was systematically varied so that half of the *S*s in each group received digits as A terms and letters as B terms, while the reverse contingency was used for the remaining *S*s. Four different orders of items occurred before the lists repeated. There was a 2-sec. interval between each trial.

The *S*s were 105 female undergraduates fulfilling psychology requirements, 20 being randomly assigned to each of the 5 groups. Five *S*s (1 each in Groups 0, 2, and 8, and 2 in Group 4) failed to reach criterion on either the A–B or B–A list within 30 trials and were excused, being replaced by the next arriving *S*s.

Results. In an analysis of total trials to criterion, no main effect for group, $F < 1$, or for order, $F < 1$, was found. It appeared as if the groups learned the lists at the same rates overall and were thus comparable. There was, however, a highly significant difference between the total trials to criterion in Stages 1 and 2, $F(1, 90) = 57.71$, $p < .001$. Since *S*s were required to

[1] Requests for reprints should be sent to C. Michael Levy, Department of Psychology, Psychology Building, University of Florida, Gainesville, Florida 32605.

learn to 1 of 5 possible criteria during A–B learning but to a criterion of 1 errorless trial during backward learning, the average number of trials to reach criterion on A–B was suppressed relative to B–A.

The degree of learning during Stage 1 had a reliable effect on performance during Stage 2, $F(4, 90) = 42.15$, $p < .001$. The mean number of trials required by the major groups to reach criterion in Stages 1 and 2 were, respectively, Group 2, 5.1 and 19.9; Group 4, 7.7 and 18.9; Group 6, 14.0 and 10.8; and Group 8, 17.9 and 7.9; Group 0 required 21.0 trials to learn the materials to 1 errorless trial. The inverse relationship between trials to criterion required for Stages 1 and 2 suggests that some B–A associations were being formed during A–B learning.

The Stage × List Order (letter–numeral vs. numeral–letter) interaction was also significant, $F(1, 90) = 6.3$, $p < .001$. The *S*s given letters as A terms and numerals as B terms performed proportionally better during A–B learning, requiring an average of 7.84 trials to criterion, than their counterparts who were presented numeral–letter combinations, requiring 9.98 trials, and proportionally worse during Stage 2 (16.86 vs. 14.52 trials, respectively). This suggests that the letter–numeral combinations were easier than the reversed pairings, reflecting either greater absolute availability for the numerals than the alphabetical characters or a higher expected value for correct responses due to guessing, since the number of potential items in the pool of single digits is less than the number of potential letters.

Prior to reaching the criterion necessary for a shift to Stage 2, Groups 4, 6, and 8 each passed one or more of the critical criteria. By combining appropriate groups during Stage 1, it was possible to determine the performance on the first postcriterion trial, a determination that was necessary for the evaluation of Trial 1 performance during B–A learning for those proceeding to Stage 2. Thus, for example, the scores of Groups 6 and 8 were averaged on the first trial after the criterion of 4 correct responses was achieved during Stage 1 in order to evaluate the B–A strength of Group 4 on the first B–A anticipation trial during Stage 2. At all comparable points there was a reliable decrement ($p < .01$) in performance averaging 28.6% at Criterion 2, 25.6% at Criterion 4, and 23.8% at Criterion 6 when *S*s were shifted from A–B to B–A learning. When only the anticipations correct on both the criterion trial and first postcriterion trial were examined, the corresponding decreases were 38.7%, 21.4%, and 24.0%. Thus, as Wollen, Allison, and Lowry (1969) have shown, the principle of associative symmetry was neither supported on an individual item-by-item basis nor in the more general case of the average number of items anticipated.

EXPERIMENT II

In Experiment I, a 1-min. delay between the critical trial in Stage 1 and the first trial of Stage 2 was necessary in order to change lists on the me[mory] drum. But the corresponding interval betwee[n the] criterion trial and the first postcriterion tria[l for] the comparison groups continuing in Stage 1 [was] only 2 sec. Additionally, in the previous c[om]parisons, there was a potential bias in favor [of] continuing A–B learning who were not subj[ect to] the loss of set. A second experiment was the[n] designed in which 3 groups were interrupted du[ring] Stage 1 learning after reaching a given criter[ion] and were engaged in conversation for 1 min. [Ac]quisition of Stage 1 was then resumed.

Method. The materials were identical to those used in Experiment I.

The *S*s were 31 female undergraduates ta[ken] from the same pool and during the same acade[mic] quarter as those who served in Experiment [I]. One *S* failed to reach the Stage 1 criterion wi[thin] 30 trials and her data were voided.

Three groups of 10 *S*s each participated [in] Stage 1 until a given criterion of learning [was] reached, after which they were engaged in co[n]versation with *E* for 1 min. Training then resu[med] for Stage 1. Group 4A was interrupted a[fter] making 2 correct responses and proceeded to [a] criterion of 4 out of 8 correct anticipati[ons]. A second delay interval then occurred du[ring] which *E* changed the lists on the memory d[rum] and after which Stage 2 began. The proced[ure] for Group 8A was similar except that a 1-mi[n.] interruption for conversation occurred after each of the first 3 criteria (2, 4, or 6 correct responses) were reached. The group was subsequently trained to a criterion of 1 perfect recitation (and give[n] a final interruption) prior to commencing Stage 2. Groups 4A and 8A learned the B–A list to a criterion of 1 errorless trial. Group 8B was trained to a criterion of 1 errorless trial in Stage 1, and was then interrupted for 1 min. of conversation before receiving 2 additional trials of Stage 1 learning. It received no training on Stage 2.

Results. Prior to determining the effects of the 1-min. interruption, the preinterruption performance of Groups 4A, 8A, and 8B was compar[ed] to Groups 4, 6, and 8 used in Experiment I. [The] 6 groups entered one analysis of variance of the number of trials to reach criteria of 2 and 4 correct anticipations in Stage 1. Groups 6, 8, 8A, and 8B were compared on the number of trials to criteria of 2, 4, and 6 correct anticipations. Data from Groups 8, 8A, and 8B were used in a third analysis of trials to reach 2, 4, 6, and 8 correct anticipations. Inasmuch as the groups main effects fell far short of significance in these analyses (all Fs < 1), the preinterruption A–B performance of the groups in Experiments I and II were pooled for subsequent comparisons.

In order to evaluate the effects of the 1-min. interruption on subsequent A–B and B–A performance, 2 post hoc groups were formed. Forward-interrupted *S*s, composed of Groups 4A, 8A, and 8B, were given 1 or more 1-min. interruptions during A–B learning. Backward-interrupted *S*s, including Groups 4, 6, and 8, had their major interruption immediately prior to B–A learning.

After each learning criterion was attained, the number of correct responses was determined on the first postinterruption trial. Table 1 indicates that there was an inverse relationship between the Stage 1 criterion and postinterruption improvement. The decrement at Criterion 8 and the increment at Criterion 2 in the number of correct anticipations after the interruption were highly reliable, yielding *z*s of 3.67 and 2.98, respectively. When *S*s were shifted to the first trial of Stage 2, their correct anticipations decreased by 25%–30% relative to *S*s who were given a 1-min. interruption before resuming Stage 1 learning. Except at Criterion 4, all of the decrements were statistically reliable.

Table 1 also presents the mean number of anticipations that were correct on both the criterion trial and the first postinterruption trial. Except for the limiting case at Criterion 8, there were fewer correct anticipations in this analysis than in the preceding one. But, as in the previous analysis, the first-trial performance on Stage 2 was significantly inferior to that after a 1-min. interruption for those continuing in Stage 1 in each comparison except Criterion 4.

Discussion. Although the methodologies used in the present study and by Baumeister and Campbell (1971) were quite different, the data are strikingly similar. When *S*s were given highly available materials, a considerable amount of backward learning occurred at each test criterion in forward learning. The A–B and B–A functions were approximately mirror images.

The present data, however, offer strong contraindications for the associative symmetry hypothesis (Asch & Ebenholtz, 1962) and for the revision of this hypothesis proposed by Wollen et al. (1969). If it may be assumed that the 1-min. interruption used in Experiment II controlled for the disruption of set and/or for the opportunities to forget in *S*s shifted to Stage 2, then the associative symmetry hypothesis remains untenable. Regardless of whether the dependent variable was the average number of postcriterion anticipations or the average number of anticipations correct on both criterion and postcriterion trials, performance on the B–A task was poorer than it was at the comparable points in A–B learning.

It is equally important to note that, even with rapid rates of presentation used to minimize the opportunities for backward rehearsal, an appreciable amount of B–A learning occurred during the process of A–B learning. The backward associative strength, after correcting for the effects of the 1-min. interruption, was estimated at 70%–75%

TABLE 1

MEAN NUMBER CORRECT ANTICIPATIONS

Stage 1 criterion	First postcriterion trial: Forward interrupted	First postcriterion trial: Backward interrupted	First postcriterion trial: p	Criterion trial *and* first postcriterion trial: Forward interrupted	Criterion trial *and* first postcriterion trial: Backward interrupted	Criterion trial *and* first postcriterion trial: p
2	2.7	1.9	<.05	1.6	.9	<.01
4	4.2[a]	2.9	*ns*	2.8[a]	2.2	*ns*
6	5.6[a]	4.1	<.05	4.4[a]	3.4	<.05
8	6.4[a]	4.9	<.01	6.4[a]	4.9	<.01

[a] The $n = 10$; otherwise, $n = 20$.

of the strength of the originally learned materials, independent of the level of A–B learning.

Obviously, some pairs were either learned bidirectionally or at least had achieved sufficient bidirectional associative strength to permit successful elicitation at the time of testing. These particular pairs were, in general, idiosyncratically determined. There was a tendency for the letter–numeral combinations to produce superior performance than the numeral–letter pairs, particularly during the initial trials of Stage 2, suggesting that (*a*) despite the high association value of the materials, the equivalence of availability of the letters and numerals was not achieved at a functional level, and/or (*b*) the expected chance value of correct responses due to guessing depended upon the class of responses required.

Given the utility of the concept of backward associations as a mechanism for understanding many phenomenon in verbal learning (see Ekstrand, 1966), and the similarities between the growth of forward and backward associations in the present experiments, future research should be directed toward investigations of the susceptibility of backward learning to the major variables in forward-learning situations.

REFERENCES

Asch, S. E., & Ebenholtz, S. M. The principle of associative symmetry. *Proceedings of the American Philosophical Society*, 1962, **106**, 135–163.

Baumeister, A. A., & Campbell, C. Formation of backward associations in paired-associates learning by normal children and retardates. *Journal of Experimental Psychology*, 1971, **89**, 298–305.

Ekstrand, B. R. Backward associations. *Psychological Bulletin*, 1966, **65**, 50–64.

Jantz, E. M., & Underwood, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *Journal of Experimental Psychology*, 1958, **56**, 174–179.

Wollen, K. A., Allison, T. S., & Lowry, D. H. Associative symmetry versus independent association. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 283–288.

(Received January 29, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 330–332

OPERANT CONDITIONING OF VASOCONSTRICTION: A VERIFICATION[1]

ROBERT M. STERN[2] AND RAYMOND P. PAVLOSKI

Pennsylvania State University

A verification of operant conditioning of unelicited vasoconstrictions was obtained for 2 separate sets of data. Significant differences were found between 30 experimental *S*s and 60 yoked control *S*s, half of whom received reinforcement identical in time and number to a given experimental *S* and half of whom received the same number of reinforcements per minute but who were never reinforced when they were vasoconstricting. Postsession questionnaire data revealed significant group differences in explanations of the response reinforcement contingencies.

Several studies have reported success in operant conditioning of vasomotor responses in animals, but only 2 published papers (Kimmel & Kimmel, 1967; Snyder & Noble, 1968) have presented evidence for operant conditioning of vasoconstriction in human *S*s; and in the Kimmel and Kimmel study, relevant statistical tests are not provided. Because of the important theoretical and methodological issues involved and the possible application of these techniques to psychosomatic disorders, it was deemed desirable to repeat the Snyder and Noble study. In the 2 experiments to follow, attention was given to possible cognitive mediators and to the role of reinforcement for the so-called noncontingent control groups (Stern, 1967).

EXPERIMENT I

Method. The *S*s were 27 female and 18 male students from introductory psychology classes at Pennsylvania State University. Right index-finger blood volume was recorded using a Beckman photoplethysmograph transducer, and respiration was recorded by means of a Parks Electronics silastic mercury strain gauge transducer. The reinforcing stimulus was a 28 v. pilot light, which was presented for 3 sec.

The *S*s were divided into 3 groups of 15 each; all were read the instructions used by Snyder and Noble (1968) except for the absence of arm electrodes and for the shortening of session length from 40 to 30 min. (no extinction period). Basically, the instructions told *S*s to relax and not try to make any voluntary responses. After the instructions were read, a 5-min. baseline period was followed by a 25-min. acquisition period, after which a questionnaire was administered. The criterion for a vasoconstriction was set independently for each *S* during the baseline period according to the procedure utilized by Snyder and Noble (1968). During this period, responses were recorded and reinforcement was omitted; the minimum amplitude response reinforced during the acquisition period was the largest baseline period vasoconstriction free from respiratory and muscular artifacts. This means that during the baseline period each *S* made 1 criterion response except for the unusual case where *S* made 2 or more "largest" responses of equal size.

Only experimental (Group 1) *S*s were reinforced for vasoconstricting during the conditioning period. Any vasoconstriction recorded within 15 sec. following a reinforcement was not reinforced. The 30 control *S*s were divided into 2 groups: each *S* in Group 2 (truly yoked) received reinforcement at the same relative time as his experimental counterpart, while each *S* in Group 3 (partially yoked) received the same number of reinforcements per minute, but never when *S* was vasoconstricting. At the end of each session, *S* was given a questionnaire to complete that contained the following queries: (*a*) What do you think made the light come on? and (*b*) If I told you that it was a response that you made that determined *when* the light came on, what do you think it was? An extinction period was not included in the sessions because of the interest in *S*s' ability to state the response reinforcement contigency. If *S*s had in fact successfully tested hypotheses about what provoked the reinforcement, an extinction period would have thoroughly confused them and invalidated the questionnaire data.

Results and discussion. Figure 1 shows the mean number of artifact-free vasoconstrictions that met the amplitude criteria for successive 5-min. periods.

The mean levels of responding for the 3 groups during the acquisition phase were 3.9, 1.8, and 0.9. Analysis of variance of the acquisition period data revealed significant group differences, $F\ (2, 42) = 12.05$, $p < .001$. Individual comparisons indicated that the level of responding of the experimental group was significantly greater than both control groups at each of the 5 time periods. The truly yoked group responded at a significantly higher level than the partially yoked group during Periods 4 and 5. Although no significant overall time effect was found, there was a significant Group $\times$ Time interaction, $F\ (7, 144) = 2.89$, $p < .01$. The latter is given with adjusted degrees of freedom (Huynh & Feldt, 1970).

Results from the postsession questionnaire data were analyzed to see whether or not the different reinforcement contingencies created group differences in regard to expectancy of reinforcement. The

[1] The second author was an NSF Undergraduate Research Participant during the course of this study.

[2] Requests for reprints should be sent to Robert M. Stern, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802.

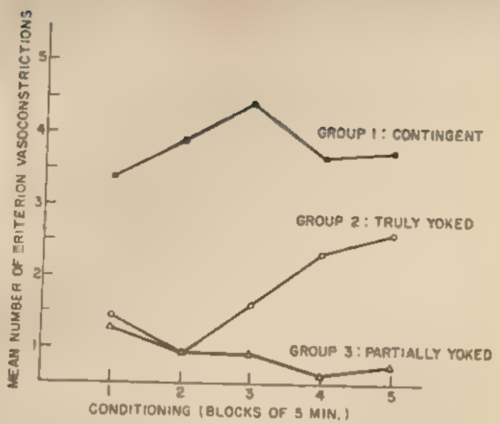


FIGURE 1. Mean number of criterion vasoconstrictions of contingent and control groups, Experiment I.

majority of *Ss* in the contingent group thought that their stimulating thoughts or bodily activity made the reinforcement light come on. Most *Ss* in both of the control groups, on the other hand, reported that the reinforcement was random or the contingency was unknown to them. The results of the questionnaire data, while not significant, were in the predicted direction. The truly yoked control group gave the expected high percentage of responses of contingency unknown. It is interesting to note that despite this reply their level of vasoconstriction increased considerably from Trial Block 2 to Trial Block 5 of the acquisition period, as shown in Figure 1. Conceivably, this increase in sympathetic responding may have been due to *Ss* persistently testing various hypotheses concerning the response reinforcement contingency.

Experiment II

The vasomotor and subjective report data of Experiment I were thought to be sufficiently interesting to repeat the experiment. More specifically, the experimental group did make significantly more vasoconstrictions than the control groups, as Snyder and Noble (1968) reported, and the questionnaire data, as mentioned above, were in the predicted direction but insignificant.

Method. The design as described above was repeated using 45 new *Ss*.

Results and discussion. Figure 2 shows the mean number of artifact-free vasoconstrictions that met the amplitude criteria for successive 5-min. periods.

The mean levels of responding for the 3 groups during the acquisition phase were 4.0, .9, and .1. Analysis of variance of the acquisition period data revealed significant group differences, F (2, 4) = 71.60, $p < .001$. Individual comparisons indicated once again that the level of responding of the experimental group was significantly greater than both control groups at each of the 5 time periods. There were no significant differences between the control groups except at Time Interval 4, where truly yoked *Ss* made more vasoconstrictions than partially yoked *Ss*. Neither the overall time effect nor the Group × Time interaction were significant.

The results of the postsession questionnaire data indicated that *Ss* in the contingent group tended to report that their stimulating thoughts or bodily activity made the light come on, whereas *Ss* in the 2 control groups reported either that they did not know what the contingency was or that reinforcement was dependent on their relaxing. These group differences were found to be highly significant, χ^2 (2) = 17.00, $p < .001$.

Conclusions

The 2 sets of data from the present experiments, collected under identical conditions, show successful modification of the rate of vasoconstrictions. This is in support of the earlier findings of Kimmel and Kimmel (1967) and Snyder and Noble (1968). Postsession questionnaire data revealed that most experimental *Ss* hypothesized correctly that the reinforcing light was contingent upon their stimulating thoughts or bodily activity, whereas most control *Ss* either thought that it was contingent on their relaxing or they did not know.

A rather curious finding which can be seen in both Figures 1 and 2 was that the performance of experimental groups reached almost the highest level during the very first 5-min. conditioning period. No ready explanation is available for this finding, even though it was observed in some of the earliest studies of operant conditioning of autonomic responses (Fowler & Kimmel, 1962).

Snyder and Noble (1968) reported a much higher level of responding than do the present authors; this may be attributed to a methodological difference. In the former study "any definite vasoconstriction not preceded by bodily movement . . . [p. 264]" was reinforced; only in the results were respiratory components and amplitude criteria taken into account. Of interest is the fact that in the present study, in which considerably fewer reinforcements were presented (since amplitude and respiratory criteria were considered during each session), results again point to successful modification of the rate of vasoconstriction. The mecha-

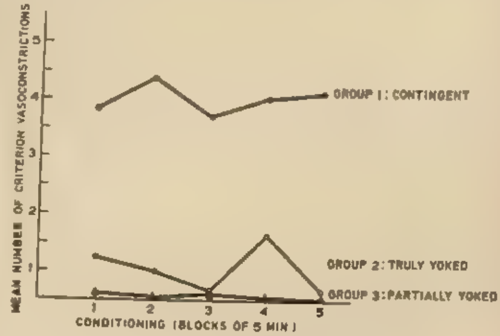


FIGURE 2. Mean number of criterion vasoconstrictions of contingent and control groups, Experiment II.

nisms by which this change comes about are not known at this time; however, an important aspect of these findings is that vasoconstriction, an autonomic response of great relevance to hypertension and other psychosomatic diseases, can be relative easily controlled and/or conditioned.

REFERENCES

Fowler, R. L., & Kimmel, H. D. Operant conditioning of the GSR. *Journal of Experimental Psychology*, 1962, **63**, 563-567.

Huynh, H., & Feldt, L. S. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association*, 1970, **65**, 1582-1589.

Kimmel, H. D., & Kimmel, E. Inter-effector influences in operant autonomic conditioning. *Psychonomic Science*, 1967, **9**, 191-192.

Snyder, C., & Noble, M. Operant conditioning of vasoconstriction. *Journal of Experimental Psychology*, 1968, 263-268.

Stern, R. M. Operant conditioning of spontaneous GSR: Negative results. *Journal of Experimental Psychology*, 1967, **75**, 128-130.

(Received February 20, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 332-334

WHOLE-LIST RETENTION FOLLOWING WHOLE-PART LEARNING [1]

RONALD OKADA[2]
Atkinson College, York University, Downsview, Ontario, Canada

STEPHEN T. CAREY[3]
University of Toronto, Toronto, Ontario, Canada

After 6 trials of whole-list learning and 5 trials of part-list learning, *S*s were asked to reproduce (Experiment I) or relearn (Experiment II) the items from the whole list. Experimental *S*s, whose part list was taken from the whole list reproduced (Experiment I) and relearned (Experiment II) more whole-list items than control *S*s, whose part list did not contain any items from the whole list. These results are incompatible with Tulving and Osler's 1967 theory of the reorganization of whole-list subjective units during part-list acquisition.

Tulving and Osler (1967) have argued that the whole-part paradigm provides data that strongly support the organizational or interdependence hypothesis of free-recall learning. In this paradigm, *S*s initially learn a list of words (whole list) under conditions of free recall. Then, experimental *S*s receive further trials on a part list that is composed of a subset of the whole-list items, while control *S*s learn a part list that does not contain items from the whole list. The critical finding is that experimental *S*s exhibit less efficient learning of the part list than do control *S*s. This counterintuitive finding, Tulving and Osler maintain, results from the fact that the formation of new subjective (S) units in the part list is hindered by the prior organization of the words in different S units in the whole list. Thus, during part-list learning, experimental *S*s "must modify the units formed during whole list learning or else form completely new units [Tulving & Osler, p. 255]." Ehrlich (1970) also favors the view that the organization of the whole list is destructured during part-list learning in order that new units can be formed; however, he maintains that the whole-list organizational structure is not permanently disrupted but only temporarily inhibited. Ehrlich reached this conclusion following the demonstration that whole-list retention was not affected by several intervening trials of second-list learning in which the second list was comprised of alternate subsets of whole-list items.

If the organization established during whole-list learning is not permanently disrupted during part-list learning, one would expect that experimental *S*s should prove as able as control *S*s in their capacity to recall the whole list following part-list learning. However, if experimental *S*s must reorganize the words in the part list, and if recall is dependent upon the retrieval of S units (Tulving & Osler, 1967), then experimental *S*s must prove inferior to control *S*s at subsequent whole-list recall. The present studies were designed to compare whole-list reproduction (Experiment I) and whole-list relearning (Experiment II) for experimental and control *S*s after each group had learned its respective part list. Since there is no apparent reason to postulate that control *S*s must reorganize their whole list during part-list learning, this condition is an appropriate control for assessing whole-list retention following part-list learning.

Method. Fifty students of both sexes at the Brant Senior Public School in Toronto served as *S*s in Experiment I. They were randomly assigned to 2

[1] This research was supported by Grant A8266 from the National Research Council of Canada to the first author and a National Research Council fellowship to the second author. The authors wish to thank Doug Smith and Bill Kirkwood for arranging for the testing of *S*s.

[2] Requests for reprints should be sent to Ronald Okada, Department of Psychology, Atkinson College, York University, Downsview, Ontario, Canada.

[3] Now at the University of Alberta, Edmonton, Alberta, Canada.

groups of 25 each with the restriction that each of the groups have an equal number of males and females. All *S*s were seventh-grade students ranging in age 12–14 yr.

Forty-eight students of both sexes at the Bramalea Secondary School in Brampton served as *S*s in Experiment II. They were randomly assigned to 2 groups of 24 with the restriction that each of the groups have an equal number of males and females.

In Experiment I, all *S*s learned a common whole list prior to the part list. Experimental *S*s learned a part list that consisted of items randomly selected from the previously learned whole list. Control *S*s learned a part list that had no words in common with the whole list. All lists consisted of 2-syllable nouns from the 1,000 most common words in the Thorndike–Lorge (1944) word count. The whole list contained 24 words; the part list contained 12 words. In Experiment II all details were the same as in Experiment I except that experimental and control *S*s learned a different whole list and a common part list, as in the Tulving and Osler (1967) study.

The experimental and control groups were treated identically during the learning of the whole and part lists in both experiments. After standard free recall instructions were read, the first whole list was presented via tape recorder at a 1-sec. rate. Ninety seconds were allowed for written recall. There were 6 alternating training and test trials with the items computer randomized on each trial. Before the first presentation of the part list, both groups were told that "we will now go on to a shorter list. Listen carefully to each word and write down the words in any order after the list has been read." As in the Tulving and Osler (1967) study, experimental *S*s were not informed about the relationship of the second list to the first list. The part list was presented via tape recorder at a 1-sec. rate. Forty-five seconds were allowed for written recall There were 5 alternating training and test trials with items computer randomized on each trial.

In Experiment I, after the test phase following the fifth presentation of the part list was completed, *S*s in both groups were given the following instructions: "You are now to write down the words from the first or longer list that you studied in this experiment. Do not write any of the words from the second or shorter list you heard. You may write these longer list words in any order, but remember, do not write the words from the second or shorter list you studied." After 2 min. of free recall of the whole list, *S*s turned to a new sheet of paper. They were then again given the instructions to recall the whole list in an attempt to determine if there was an interaction between final whole-list recall trials and conditions. The *S*s were given 2 min. for their second recall phase

In Experiment II, after the test phase following the fifth presentation of the part list was completed, *S*s in both groups were given the following instructions: "You are now going to hear the words from the first or longer list that you studied in this experiment. Listen carefully to each word and write down the words in any order after the list has been read." Ninety seconds were allowed for written recall.

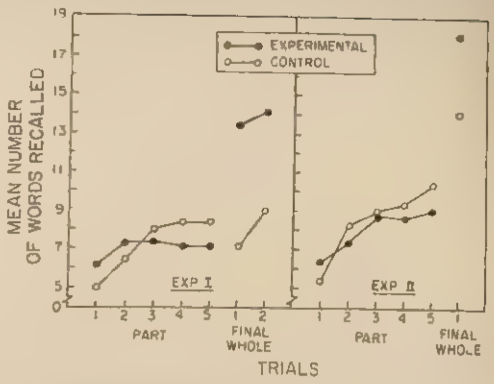


FIGURE 1. Mean number of words recalled during part-list acquisition and final whole-list reproduction (Experiment I) and relearning (Experiment II).

Results and discussion. The whole-list learning data were virtually identical for both conditions in Experiment I. An analysis of variance revealed that only trials was significant, F (5, 240) = 115.17, $p < .01$. On the sixth trial of the whole list, experimental *S*s recalled 12.7 items, while control *S*s recalled 12.4 items, F (1, 48) < 1, *ns*.

The part list data are presented in the left panel of Figure 1. Both the trials effect, F (4, 192) = 32.41, $p < .01$, and the Condition × Trials interaction, F (4, 192) = 9.86, $p < .01$, were significant.

An analysis of the final whole-list reproduction data (see left panel of Figure 1) revealed that conditions and trials were significant, F (1, 48) = 37.70, $p < .01$, and F (1, 48) = 6.11, $p < .05$, respectively. However, the Condition × Trial interaction was not significant, F (1, 48) = 3.44, *ns*.

The performance of experimental and control *S*s did not differ during whole-list learning; however, during part-list learning, experimental *S*s exhibited less efficient learning than control *S*s. These results are consistent with previous studies (Tulving & Osler, 1967). With regard to final whole-list recall, Figure 1 shows that the ability of experimental *S*s to ultimately reproduce whole-list items was not impaired by several trials of part-list learning and was greater than that of the control group. These findings are inconsistent with the hypothesis that experimental *S*s must reorganize existing S units acquired during whole-list learning and form new S units in order to learn the part list. However, it might be argued that the inferiority of control *S*s in reproducing the whole list was due to the inaccessibility of the whole-list structure as a result of *S*s' recent preoccupation with an unrelated part and consequent response set suppression of the whole list.

The results of Experiment II are relevant to this point. In this study, a relearning trial on the whole list followed whole–part learning to determine whether a single reinstatement of the whole-list context would benefit control *S*s in retrieving items

from the whole list. The whole-list learning data were virtually identical for both groups in Experiment II. An analysis of variance revealed the only significant effect to be that of trials, F (5, 230) = 137.75, $p < .01$. On the sixth trial of the whole list, experimental *S*s recalled 16.5 items, while control *S*s recalled 16.9 items, F (1, 46) < 1, *ns*.

The part-list data are presented in the right panel of Figure 1. Both the trials effect, F (4, 184) = 80.32, $p < .01$, and the Condition × Trials interaction, F (4, 184) = 7.29, $p < .01$, were significant.

On the relearning trial of the whole list (see the right panel of Figure 1), experimental *S*s recalled 18.4 items, while control *S*s recalled 14.3 items, F (1, 46) = 13.93, $p < .01$.

Again as in Experiment I, the typical negative transfer effects were found during second-list learning. Nevertheless, on the subsequent relearning trial of the whole list, experimental *S*s recalled more whole-list items than control *S*s. Thus, the additional presentation of the whole list does not alter the fact that experimental *S*s recall more whole-list items than do control *S*s. For the reasons stated earlier, this finding is contrary to the predictions from Tulving and Osler's (1967) interdependence hypothesis, which states that whole-list S units which mediate recall are modified or completely new S units are formed during part-list learning. If new S units are organized during part-list acquisition, why then do experimental *S*s show no loss of whole-list items in subsequent whole-list recall and achieve higher recall performance than control *S*s, who had no reason to modify their S units during part-list learning?

Slamecka, Moore, and Carey (1972) offered an alternative account of negative transfer in the whole–part paradigm. This account proposes that the greater discrimination that experimental *S*s are required to perform between first and second lists during part-list learning is primarily responsible for the negative transfer. Inability to discriminate among part-list items and whole-list items would depress part-list performance, producing negative transfer during part-list learning.

A list-discrimination hypothesis could also accommodate the reproduction and relearning data following whole–part learning. Since ex *S*s have an opportunity to retrieve whole-l… items during part-list learning, discriminati… them for part-list items on the basis of … of occurrence, these *S*s should be adep… ducing or relearning the whole list followi… part learning. Control *S*s, in learning a n… items during part-list learning, have no r… retrieve whole-list items and are not hind… the list-discrimination task (and thus acqui… part list more readily). However, when recall of th… whole list is required following whole–part learning, the frequency information that was established and utilized during whole-list learning (situational frequency) has begun to assimilate to its preexperimental level (background frequency) during the interpolated part-list learning (see Underwood & Freund, 1970). Thus, control *S*s exhibit forgetting of whole-list members following part-list learning.

In summary, the superiority of experimental *S*s at reproducing (Experiment I) and … (Experiment II) the items from the w… following whole–part learning may derive from … more frequent retrieval of whole-list items for experimental *S*s. In any event, since expe… *S*s retain more of the whole list following part-li… learning than do control *S*s, it cannot be the case that the organizational structure that mediates retrieval is permanently disrupted. Similar conclusions were reached in the part–whole paradigm by Carey and Okada (1973).

REFERENCES

CAREY, S. T., & OKADA, R. Part-list recall following part/whole learning. *Memory and Cognition*, 1973, 1, 172–176.

EHRLICH, S. Structuration and destructuration of responses in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 282–286.

SLAMECKA, N. J., MOORE, T., & CAREY, S. T. Part-to-whole transfer and its relation to organization theory. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 73–82

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbo… 30,000 words*. New York: Columbia University, Te… College, Bureau of Publications, 1944.

TULVING, E., & OSLER, S. Transfer effects in whole-part f… recall learning. *Canadian Journal of Psychology*, 1967, 2… 253–262.

UNDERWOOD, B. J., & FREUND, J. S. Retention of … crimination. *Journal of Experimental Psychology*, … 1–14.

(Received February 20, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 335-338

COMMON VERBAL QUANTIFIERS: USAGE AND INTERPRETATION[1]

MARILYN A. BORGES
San Diego State University

BARBARA K. SAWYERS[2]
University of California, San Diego

The *S*s made judgments about the meanings of 6 common quantifiers combined with different set sizes. Experiment I involved nonverbal quantity judgments; Experiment II, verbal numerosity judgments. An information integration analysis showed that *S*s' judgments followed a simple multiplying rule: the meanings of the quantifiers increased linearly with respect to the set size of the population considered. The *S*s used the quantifiers in a highly consistent manner, the only exception being the quantifier *several*, which was found to have 2 distinct usages.

The present experiments studied the usage of 6 quantifiers: *few*, *several*, *some*, *lots*, *many*, and *most*. These quantifiers are all words occurring with high frequency in the English language (AA count, Thorndike & Lorge, 1944). They are independent from one another in etymological origins (there are no common cognates or roots), and their meanings are not precisely defined in a numerical sense, e.g., *few* is more than 2, but not *many* (Onions, 1955). In addition, the use of these quantifiers is, in most cases, learned pragmatically rather than pedagogically.

Because these quantifiers are frequently used, a high degree of agreement about their meanings should exist. The main experimental questions to be answered were as follows. (*a*) Does an individual use these quantifiers in a consistent manner? (*b*) What is the extent of agreement among adults on the uses of these general quantifiers? (*c*) How do *S*s modify their interpretations of each quantifier depending upon the size of the set on which they operate?

The *S*s' tasks were to make judgments about the use of each quantifier. Judgments were based on 2 sources of information: (*a*) *S*'s conception or understanding of the meanings of the quantifiers and (*b*) *S*'s perception of the size or numerosity of the sets from which he chose. In Experiment I, values for each quantifier were determined on the basis of quantity, i.e., *S* formed subsets from various set sizes of marbles to represent each quantifier. In Experiment II, each judgment was a cognitive rating of "count" or numerosity, with the set information supplied to *S* and the quantifier response given by *S* in the form of specific integers.

Previous investigators have used statistical multiplying models for evaluating word combinations. Cliff (1959), extended by Howe (1962), studied combinations of adverbs and adjectives. However, this previous work has a serious shortcoming (Anderson, 1973) because it depends on Thurstonian scaling methods. For verbal stimuli, this method relies completely on pooled *S* data, making no allowance for individual differences. Integration theory, which takes individual differences into account, has been used in a variety of judgmental tasks (Anderson, in press; Sawyers & Anderson, 1971). The present study applied this general analytic approach to determine how *S*s integrate 2 diverse sources of information, quantifier meaning and set size, in making their judgments.

METHOD

The *S*'s task was to make a series of judgments about the quantitative meanings of 6 different quantifiers combined with 4 (Experiment I) or 5 (Experiment II) different set sizes. In Experiment I, judgments were made by taking marbles from one box and placing them in another box. In Experiment II, *S*s made judgments by specifying a particular integer for each combination of quantifier name and numerical set size.

Subjects. The *S*s were 56 students (26 in Experiment I and 30 in Experiment II) from San Diego State University.

Design and Procedure

Experiment I. There were 72 judgments for each *S*, with 6 quantifiers and 4 set sizes. The quantifiers were *few*, *several*, *some*, *many*, *lots*, and *most*. Sets of Size 12, 24, 36, and 48 were used. Each *S* served in 3 replications of this design. For each replication, *S* judged all the quantifiers in random order in 1 set size for each of the 4 set sizes. Within each replication, set size was presented in a different random order for each *S*.

The *S*s were run individually and instructed to make judgments by taking marbles from one box and putting them into another box, without actually counting the marbles. The *S*s were allowed to take as many handfuls of marbles as necessary and could rearrange the marbles between the 2 boxes until they were satisfied that their judgments were accurately reflected. Although no time limit

[1] This research was supported in part by National Science Foundation Grant GB-20798 and U.S. Public Health Service Grant MH-15828 from the National Institute of Mental Health. Responsibility for authorship is equally shared. The authors wish to thank Norman H. Anderson and George Mandler for their invaluable assistance.

[2] Requests for reprints should be sent to Barbara K. Sawyers, Department of Psychology, University of California, San Diego, La Jolla, California 92037.

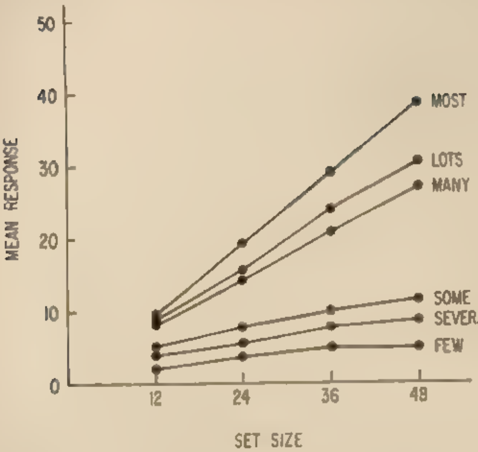


FIGURE 1. Mean numerical judgment of the 6 quantifiers for Experiment I.

was imposed, it was suggested that *S* work at a fairly rapid pace. Before the experiment began, *S*s were shown a list of the 6 quantifiers. However, *S*s were not informed of the set sizes to be used in the experiment. Two 9 × 16 × 7.5 mm. wooden boxes were placed in front of *S*. One of the boxes contained 1.5-mm. marbles, of the first set size to be judged. For each trial, a statement of the following type was read to *S* by *E*: "Take *some* of the marbles from this box [*E* pointed to first box] and put them into this box [*E* pointed to second box]." After each judgment, *E* recorded the number of marbles transferred by *S* to the second box, then returned the marbles to the first box for the next judgment. After each series of 6 judgments on one set size, *E* changed the number of marbles in the first box, and the next series of judgments was made.

Experiment II. Each *S* made a total of 90 judgments. The quantifiers used were those of Experiment I. Sets of Size 12, 36, 60, 84, and 108 were used. Each *S* judged the 30 quantifier-set-size combinations over 3 replications. As in Experiment I, for each replication, all the quantifiers were presented together in random order, for 1 set size at a time. Set size was presented in a different random order within each replication for each *S*.

The *S*s were run in a group and were instructed to specify a numerical value for each quantifier-set-size combination. The quantifiers to be used in this experiment were shown to *S*s before the experiment began; however, set sizes were not disclosed. Each *S* was given 4 packets, with 5 booklets in each packet. Each booklet contained 6 pages, with a statement of the following type on each page: "If you were to take *some* of the marbles from 84 marbles, you would take ——— marbles." The *S* filled in the blank with his numerical judgment. Only specific integers were allowed as responses. The first packet was given as practice to familiarize *S*s with the task. The *S*s were cautioned not to refer back to past judgments or change a judgment on a page that they had already turned.

RESULTS

Pilot work suggested that *S*s were relatively consistent in their usage of 5 of the quantifiers. For *several*, however, there were large individual differences, and *S*s could clearly be partitioned into a "low" group and a "high" group. Based on these preliminary results, *several* was assumed to have 2 distinct meanings, and *S*s were subdivided into a high-*several* or low-*several* category. In Experiment I, only 2 of 26 *S*s were in the high-*several* category; for simplicity, their data are not included. In Experiment II, *S*s were equally divided between the 2 groups.

Main Empirical Results

Experiment I. Figure 1 shows the mean judgments of the 6 quantifiers for the 4 different set sizes for 24 *S*s. The relative ordering of the judgments for the 6 quantifiers from low to high was *few*, *several*, *some*, *many*, *lots*, and *most*. As can be seen in Figure 1, the judgments of the quantifiers increase as near linear functions of the objective set size. A Newman-Keuls test on all ordered pairs of means for each of the quantifiers showed each quantifier to have a unique semantic meaning ($p < .01$).

Experiment II. Figure 2 shows the mean judgments of the 6 quantifiers for the 5 different set sizes. Two curves are shown for *several*: low *several*, the mean of 15 *S*s who consistently gave low estimates of this quantifier (9%–25% of the

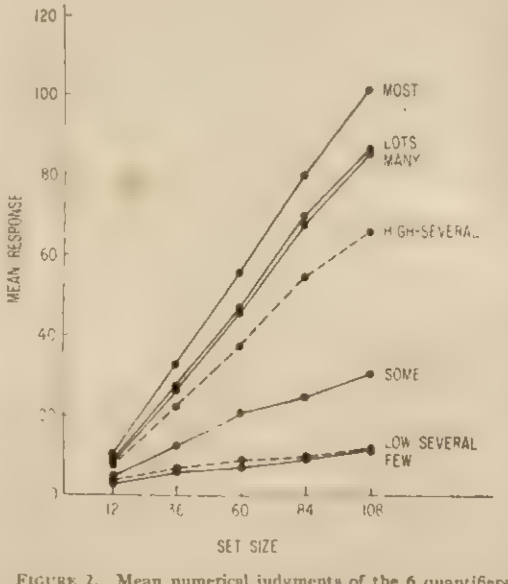


FIGURE 2. Mean numerical judgments of the 6 quantifiers for Experiment II. (For *few*, *some*, *many*, *lots*, and *most*, $n = 30$; for low *several* and high *several*, $n = 15$.)

set size), and high *several*, the mean of 15 *S*s who consistently gave high estimates (49%–77%). Newman–Keuls tests showed both high- and low-*several* *S*s to judge *lots* and *many* as having similar semantic meanings; also, low-*several* *S*s did not differentiate between *few* and *several*. However, all other quantifiers were judged as semantically unique in meaning. The ordering of the 6 quantifiers for low-*several* *S*s was *few/several*, *some*, *many/lots*, and *most*. For high-*several* *S*s, the ordering was *few*, *some*, *several*, *many/lots*, and *most*.

Model Analysis

Goodness of fit. The multiplying model from integration theory makes a simple graphical prediction: the curves of Figures 1 and 2 should form fans of straight lines, diverging as a function of increasing set size. Inspection of the curves indicates that this prediction is approximately correct.

An exact statistical test of the model is also available. Theoretically, 2 or more variables that combine in a multiplicative manner will show an interaction concentrated in the bilinear (Linear × Linear) component, but the residual should be nonsignificant. A test on the interaction residual is primary, therefore, in that a significant residual is direct evidence against the model. In Experiment I, the overall interaction (Set Size × Quantifier) was found to be very large, F (15, 345) = 101.36, $p < .01$. The residual interaction was found to be nonsignificant, F (14, 345) = 1.08. The overall Set Size × Quantifier interactions of Experiment II for both the low-*several* and high-*several* groups were also large, F (20, 280) = 203.52, and F (20, 280) = 75.19, respectively, both ps $< .01$. The residual interaction of the low-*several* group was only marginally significant, F (19, 280) = 1.70, $p = .05$. The residual interaction of the high-*several* group was nonsignificant, F (19, 280) = .47. Thus, the data give considerable support to the multiplying model: each quantifier is integrated with set size by a simple multiplying rule.

Functional scale values. Since the marbles were given as a heap in the box, it was not possible to assume without evidence that the actual objective number of marbles in the box was equal to the effective subjective number of marbles that *S* used in making his response. Thus, factors of visibility or manual accessibility might produce differences between the subjective and objective values. The functional measurement analysis from integration theory provides subjective values for both stimulus variables. This analysis showed that, for set size, subjective number was nearly a perfectly linear function of actual number. The functional scale values of the quantifiers *few*, *several*, *some*, *many*, *lots*, and *most*, respectively, were, for Experiment I, .16, .26, .33, .60, .68, and .81; and for Experiment II, .15, .17 (low *several*) and .63 (high *several*), .34, .77, .79, and .92.

Discussion

In both Experiments I and II, the multiplicative model of integration theory was found to be highly descriptive of the way in which *S*s integrate the meanings of each quantifier with set-size information. A high level of consistency was found in the use of verbal quantifiers for each individual *S* and, with the exception of *several*, across *S*s as well. There were also strong indications that *S*s judged each of the quantifiers to represent a consistent proportion of the total set size. This proportional relationship existed regardless of whether the task was a quantity judgment task (Experiment I), or a verbal numerosity judgment (Experiment II).

That *several* was found to have 2 distinct meanings was quite unexpected. A number of *S*s were interviewed, and each stated his strong belief that his particular use of *several* was the only "correct" usage, but no clues were obtained to explain this difference. However, the data show that both low-*several* and high-*several* *S*s use the word consistently in a variety of contexts. Specifically, high-*several* *S*s always judged *several* to mean more than low-*several* *S*s, regardless of context.

In Experiment I, all 6 quantifiers were judged to have distinctly different meanings. However, in Experiment II, the meanings of low *several* and *few* did not significantly differ, nor did the judgments of *lots* and *many*. The most likely explanation for this difference lies in the nature of the tasks. In Experiment I, *S*s made an intuitive judgment of relative quantity for each quantifier based on the perceptual mass of the total set of marbles. In Experiment II, *S*s had to give an exact number for each quantifier based on only verbal information about the total set size. Presumably, this was more difficult in that (*a*) usually, no specific number or percentage is associated with quantifiers, and (*b*) such precise numerosity judgments fall outside *S*s' normal range of experience with the quantifiers. It is possible that the necessity of making numerosity responses to the quantifiers rather than quantity judgments produced the less specific distinctions among the quantifiers.

In summary, adult *S*s were found to use the quantifiers in a highly consistent manner, with the exception of the quantifier *several*, which had 2 distinct interpretations. Information relating quantifier and set size was seen to follow a simple multiplying rule, i.e., usage of these quantifiers seems to be determined by each quantifier's proportional relationship to the total set size. While small variations in the meanings of words across people is to be expected, the surprising bimodal distribution of the meaning of *several* raises an important question about the objectivity of word meanings.

References

Anderson, N. H. Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception*. New York: Academic Press, 1973.

ANDERSON, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology.* Vol. 1. San Francisco: Freeman, In press.
CLIFF, N. Adverbs as multipliers. *Psychological Review*, 1959, 66, 27–44.
HOWE, E. S. Probabilistic adverbial quantification of adjectives. *Journal of Verbal Learning and Verbal Behavior*, 1962, 1, 225–242.
ONIONS, C. T. (Ed.) *Oxford universal dictionary.* London: Oxford University Press, 1955.
SAWYERS, B. K., & ANDERSON, N. H. Test of integration theory in attitude change. *Journal of Personality and Social Psychology*, 1971, 18, 230–233.
THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

(Received March 16, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 338–340

SHORT-TERM MOTOR MEMORY AS A FUNCTION OF FEEDBACK AND INTERPOLATED ACTIVITY [1]

L. BURWITZ [2]
University of Illinois

A test was made of the hypothesis that short-term motor forgetting could be found on 1 trial, in the absence of interference, if covert rehearsal were sufficiently restricted. The covert-rehearsal process was impaired by either reducing response-produced feedback and/or requiring attention toward a rehearsal-preventing activity during the retention interval. The task used was linear lever positioning. Two conditions of feedback and 3 levels of interpolated activity gave 6 independent treatments, with 19 *S*s each. Feedback involved the presence of audition, proprioception, and vision (APV) or their absence (−APV). Interpolated activity was a 5-sec. rest or immediate recall, a 90-sec. rest, and a 90-sec. digit classification. Under −APV, there was forgetting over the unfilled retention interval and no increased forgetting with interpolated activity. Under APV, there was no forgetting at rest, but recall decrements occurred following interpolated activity. The results were interpreted in terms of trace decay theory, and the importance of covert rehearsal for motor retention was discussed.

The trace decay theory of forgetting has received little experimental attention in short-term motor memory research. Only 2 studies have used a single-trial learn–recall design, which eliminates the possibility of interference. In these studies, Williams (1971) reported no forgetting in the absence of prior learning, while Adams, Marshall, and Goetz (1972) found evidence of trace decay under minimal feedback conditions. Adams et al. suggested that impoverished feedback produced a weak image or perceptual trace and, although they made no attempt to measure the strength of the trace, they concluded that the weak trace decayed over time. According to Brown (1958), decay is not a function of time per se, but it is rather a consequence of covert rehearsal restrictions. Thus, impoverished feedback appears to impair covert rehearsal.

If one assumes that *S* is unable to rehearse a weak perceptual trace, but that he is able to rehearse a stronger image, one would expect on the basis of the trace theory that interpolated rehearsal-preventing activity would differentially effect recall under impoverished and augmented feedback conditions. In support of this contention, Posner and Konick (1966) and Posner (1967) found that blind movements (minimal feedback) were forgotten over an unfilled retention interval and unaffected by interpolated activity, while visually guided movements showed forgetting only after rehearsal-preventing activity. Posner and Konick and Posner used a repeated measures design (48 trials in 50 min.); consequently, their results do not provide conclusive evidence of trace decay, since such designs contain a built-in potential for proactive interference (cf. Adams et al., 1972). In the present study, a single-trial learn–recall design was employed to eliminate the possibility of interference and to provide a clean test of the trace decay theory, which predicts forgetting under impoverished feedback and interpolated noninterfering activity because of covert-rehearsal restrictions. In addition, the strength of the perceptual trace was assessed under impoverished and augmented feedback conditions.

Method. One hundred and fourteen right-handed male undergraduate students were randomly assigned to 1 of the 6 independent conditions of the 2×3 factorial design, with 19 *S*s per cell. The 2 levels of the first factor involved the presence of audition, proprioception, and vision (APV) or their absence (−APV), and the second factor was

[1] The author wishes to thank Jack A. Adams for his constructive comments.
[2] Requests for reprints should be sent to L. Burwitz, who is now at the Department of Physical Education and Recreation, Acadia University, Wolfville, Nova Scotia, Canada.

a 5-sec. rest or immediate recall, a 90-sec. rest, and a 90-sec. digit classification.

Linear lever positioning was used as the memory task, and the apparatus was essentially the same as that described by Adams et al. (1972). On top of the lever, which ran along a 15-in. trackway, was a button that *S* depressed to move the carriage and released to lock it in place for the purpose of measurement. The apparatus was located behind a 1-way mirror and *E* was able to manipulate visual feedback by altering the lighting conditions on either side of the mirror. All lever movements were from right to left and were essentially frictionless (−APV) or were resisted by a coiled clock spring with a pull of 1.75 lb. (APV). Auditory feedback was eliminated when necessary (−APV) by white noise played through earphones at 120 db.

The experiment was conducted in 2 stages, the first of which involved instructions and a 20-sec. practice trial on the digit classification task. The digits were printed on a sheet of paper located on the wall to the left of *S*'s chair at approximately eye level. On the command "digit," *S* turned his head and began reading down the first of 4 columns of 2-digit numbers, verbally indicating as fast as possible whether each was high or low (above or below 50) and odd or even.

The second phase included linear-lever-positioning instructions, a demonstration (14 in.) and 2 movements, which were separated by a retention interval of 5 or 90 sec. The first movement was to the criterion (8 in.), which was defined by a stop, and the second was to where *S* thought the stop had been. At the end of each movement, *S* released the button on top of the lever and maintained his grasp for another 2 sec. before being instructed to place his right hand on the pad located on the table in front of him. The 2 movements were made under identical feedback conditions, with vision and white noise being manipulated for a period that extended from 2 sec. prior to movement initiation through 2 sec. after button release. The retention interval was timed from button release, and during this period, digit classification *S*s performed the interpolated task, while rest *S*s were required to face the mirror. All *S*s kept their right hands on the pad throughout the retention interval, and *E* returned the lever to the start.

After the 2 movements, *S* was asked to estimate the direction and magnitude of his error from the criterion. The absolute difference between the actual and estimated algebraic error scores was taken as an indication of *S*'s ability to accurately estimate his own performance. This score was assumed to reflect the strength of the perceptual trace since *S* is unable to accurately estimate performance without a strong trace. Finally, an interview was conducted to gather information on the strategy *S* used to position the lever.

Results. Analysis of variance of the absolute error scores found the main effects of both feedback and interpolated activity to be significant: $F(1, 108) = 12.55$, $p < .001$; and $F(2, 108) = 4.27$, $p = .02$, respectively. Table 1 indicates the overall superiority of the APV condition (4.9) over the −APV treatment (9.1). Post hoc analysis, using Duncan's new multiple range test, revealed that the interpolated digit classification condition (9.1) recalled significantly ($p < .05$) less than the immediate recall group (4.9), and that there were no other significant differences between the interpolated activities.

TABLE 1

Mean Absolute Retention Test Error (⅛-in. units)

Interpolated activity	Feedback condition		Grand *M*
	APV	−APV	
None (immediate recall)	3.63	6.16	4.90
90-sec. rest	3.42	10.53	6.97
90-sec. digit classification	7.68	10.47	9.08
Grand *M*	4.91	9.05	

Note. Abbreviations: APV indicates the presence of audition, proprioception, and vision; −APV indicates their absence.

The Feedback Condition × Interpolated Activity interaction was not significant, $F(2, 108) = 1.61$, $p = .20$. Although the interaction was not significant, it was expected on the basis of the trace decay theory that recall decrements would occur over an unfilled retention interval under −APV, and that forgetting would occur only after interpolated rehearsal-preventing activity under APV. To test these theoretically based a priori assumptions, 2 orthogonal tests were made, using the t ratio with the mean square error from the analysis of variance as the error term to give 108 *df*. The first test indicated that under −APV, decrements in recall were significantly ($p < .05$) larger after the 90-sec. rest (10.5) than at immediate recall (6.2). The second test revealed that under APV, there was significantly ($p < .05$) more forgetting after the 90-sec. digit classification (7.7) than after the 90-sec. rest (3.4).

The same analysis was run on algebraic error as on absolute error, and none of the tests were significant. Averaging over all interpolated activities, the mean algebraic error was −1.0 for APV and −2.9 for −APV. Averaging over feedback conditions, the means for interpolated activities were −1.1, −2.8, and −2.1 for the 90-sec. digit classification, 90-sec. rest, and 5-sec. rest, respectively.

The analysis of variance of *S*'s ability to accurately estimate his own performance (absolute difference between actual and estimated algebraic error scores) found both main effects significant: for feedback, $F(1, 108) = 7.71$, $p = .01$; and for interpolated activity, $F(2, 108) = 4.57$, $p = .01$. Table 2 indicates that −APV produced a weak perceptual trace (7.6) compared to that developed under APV (4.7). Furthermore, post hoc analysis revealed that the perceptual trace was significantly ($p < .05$) weaker following interpolated activity

TABLE 2

Absolute Difference (⅛-in. units) between Actual and Estimated Algebraic Error Scores

Interpolated activity	Feedback condition		Grand *M*
	APV	−APV	
None (immediate recall)	2.95	6.00	4.47
90-sec. rest	4.37	6.90	5.63
90-sec. digit classification	6.68	9.84	8.26
Grand *M*	4.67	7.58	

Note. Abbreviations: APV indicates the presence of audition, proprioception, and vision; −APV indicates their absence.

(8.3) than at immediate recall (4.8). None of the other differences were significant.

Discussion. The present results lend credence to the trace decay theory of forgetting, since recall decrements occured with 1 learn–recall trial in the absence of apparent interference. In support of Adams et al. (1972), there was no forgetting over the unfilled retention interval when feedback channels were intact (APV), but significant decrements in recall were induced by attenuating feedback (—APV). The manipulation of feedback also influenced *S*s ability to accurately estimate his own performance. This ability was assumed to reflect the strength of the perceptual trace, and, as Adams et al. hypothesized, the trace under impoverished feedback was significantly weaker than under APV. Apparently, minimal feedback produced a weak perceptual trace, which decayed over the unfilled retention interval because it was not rehearsed.

If one assumes that covert rehearsal was not possible following −APV but was possible after APV, then one would predict on the basis of the trace decay theory that interpolated rehearsal-preventing activity would produce no more forgetting than would rest under −APV, but that it would cause increased forgetting under APV. The results, which provided support for the hypothesis and additional evidence of trace decay, were in accord with the findings of Posner and Konick (1966) and Posner (1967) and although these studies, because of their built-in potential for proactive interference, do not offer direct evidence of trace decay, they certainly do not contradict the theory.

Short-term motor memory is apparently influenced by covert-rehearsal opportunities. The mechanism for such rehearsal is not clear at present, although it appears plausible to suggest, on the basis of the introspective reports of *S*s, that verbal labels or imaginal mediators are necessary for motor retention. The present interview data concurred with the findings of Posner and Konick (1966) and favored the imagery position, but additional research is required to substantiate this visual encoding hypothesis.

REFERENCES

Adams, J. A., Marshall, P. H., & Goetz, E. T. Response feedback and short-term motor retention. *Journal of Experimental Psychology*, 1972, **92**, 92–95.

Brown, J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

Posner, M. I. Characteristics of visual and kinesthetic memory codes. *Journal of Experimental Psychology*, 1967, **75**, 103–107.

Posner, M. I., & Konick, A. F. Short-term retention of visual and kinesthetic information. *Organizational Behavior and Human Performance*, 1966, **1**, 71–86.

Williams, I. D. The effects of practice and prior learning on motor memory. *Journal of Motor Behavior*, 1971, **3**, 205–212.

(Received March 19, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 341–343

ABSTRACTION OF UNIDIMENSIONAL CONCEPTS FROM LARGER CONCEPTUAL SYSTEMS[1]

VITO MODIGLIANI[2]

Wesleyan University

Stimuli consisting of 2 systems, a face and a pipe, were used. Each system consisted of 3 bivalued dimensions and 2 constant values. Together, the 2 systems made up a meaningful figure, a "face smoking a pipe." During an initial training phase, one of the dimensions was made relevant in a concept identification type task. In the subsequent concept conservation task, 48 *S*s were found to conserve less in response to transformations operating on the system containing the relevant attribute than to identical types of transformations operating on the other system. The findings were interpreted as indicating the importance of stimulus structure in determining the abstraction of unidimensional concepts from larger conceptual systems.

Recently, Modigliani (1971) studied how bounded- s affected the conservation of simple unidimen- nal concepts. In that study bounded stimuli nsisting of 4 bivalued dimensions were potted wers, each made up of flower pot, stem, leaves, d blossom. Unbounded stimuli were obtained rom the bounded ones by redrawing the latter a disjointed manner. Concept conservation was ower with bounded than with unbounded stimuli. he problem with this manipulation of bounded- ess is that it is not clear to what extent the effect was determined by the perceptual nature of the stimulus manipulation, rather than by conceptual boundedness per se.

In the present study, boundedness was manipu- lted both at the stimulus level, as in Modigliani 1971), and conceptually, as follows. Consider drawing depicting a "face smoking a pipe." h a stimulus is an overall bounded set of at- utes and could be perceptually unbounded ac- ding to the Modigliani procedure. On the other hand, the stimulus could also be considered as consisting of 2 separate systems, the face and the pipe, each of which constitutes a bounded set of attributes in itself. If this were so, it could be expected that a higher degree of boundedness obtained between 2 stimulus elements belonging to the same system, e.g., the pipe, than between 2 elements belonging to different systems, e.g., one to the pipe, the other to the face. Thus, the hypothesis tested in the present study was that if the concept to be conserved was a pipe element, then transformations operating on other pipe elements should be more disrupting (lead to less conservation) than similar transformations oper- ing on face elements, and vice versa. This hypothesis will be referred to as the conceptual boundedness hypothesis because it depends solely on how *S* interprets the stimulus, not on any perceptual manipulation of it.

Method. The design of the experiment was a $3 \times 2 \times 18$ factorial, the independent variables being age, perceptual boundedness, and transformations, respectively. The 3 age levels included children from the first, third, and sixth grades. The 2 perceptual boundedness levels were those used by Modigliani (1971), stimuli being either bounded or unbounded, and the transformations are described below. The first 2 factors were between-*S*s and the third within-*S*s variables.

The *S*s were 48 children, 16 from each of the 3 grades, drawn from 2 schools in Middletown, Connecticut. Within each age, *S*s were randomly assigned to the experimental conditions.

The experiment followed the same pattern as that of Modigliani (1971) and consisted of a training phase followed by a conservation test phase. In the training phase, *S*s learned to categorize stimuli according to a single relevant dimension—the shape of the eyebrows for half of them and the brightness of the pipe bowl for the other half. The training stimuli consisted of 32 outline figures drawn on 21.7×28 cm. sheets of white paper. Each drawing consisted of 10 values, of which 4 were constant, i.e., the same for all stimuli, and the remaining 6 were values from bivalued dimensions. Each of the 2 stimulus systems, the face and the pipe, consisted of 3 (varying) dimensions and 2 constant values. The 3 face dimensions were the shape of the eyebrows, the shape of the nose, and the shape of the mouth. The shape of the eyes and the face contour were constant face values. The 3 pipe dimensions were the brightness of the pipe stem, the brightness of the bowl, and the shape of the smoke the pipe emitted. The constant dimensions of this unit were the overall shape of the pipe and the shape of the bowl rim. Values were drawn at fixed positions in both conditions of perceptual boundedness. The 32 training stimuli comprised a random sample of the 64 possible stimuli, with the restriction that half contained one relevant value and the other half, the other relevant value.

The conservation phase stimuli consisted of 28 drawings, 10 of which were randomly selected from those used in the training phase, and the rest representing 1 transformation each. Trans-

[1] This research was made possible by a Wesleyan University grant in support of scholarship. The author wishes to thank Marjory S. Daltry, Clem M. Green, Harold Kaplan, and Anthony LaBella for their collaboration in providing *S*s for this research, and Robert S. Feldman for running the *S*s.

[2] Requests for reprints should be sent to Vito Modigliani, Department of Psychology, Wesleyan University, Middletown, Connecticut 06457.

TABLE 1

Mean Percentage of Conserving Responses as a Function of *Other* versus *Same* System Transformations and Perceptual Boundedness

System operated on by transformations	Perceptual boundedness	
	Unbounded	Bounded
Other	79.2	75.0
Same	80.0	53.3

formations were obtained by adding, deleting, and substituting new values for old ones in the training stimuli. For present purposes, only the first 10 transformations will be described in detail. Five of them operated on elements belonging to the system not containing the relevant attribute, referred to as the *other* system. The 5 remaining operated on elements belonging to the system containing the relevant attribute, the *same* system. For each transformation in the *other* system group there was an exact replica in the *same* system group, with 1 exception. The 5 transformations in the *other* system group were: (*a*) deletion of 1 irrelevant value, (*b*) deletion of 2 irrelevant values, (*c*) deletion of the entire *other* system, (*d*) deletion of the system's 2 constants values, and (*e*) substitution of a new value for one of the old constant ones. The same transformations were used to operate on the *same* system, the exception being that instead of deleting the entire *same* system, the relevant value was never deleted.

A test of the conceptual boundedness hypothesis entailed a comparison between these 2 groups of transformations as a function of perceptual boundedness. When the stimuli are perceptually bounded, the distinction between *other* and *same* systems is meaningful, but is effectively lost when the stimuli are perceptually unbounded. The conceptual boundedness hypothesis, therefore, predicts an interaction between perceptual boundedness and transformations operating on the *other* vs. the *same* system: there would be a significant difference due to these 2 types of transformations for bounded but not for unbounded stimuli.

Two random sequences of the 28 conservation test cards were used with the restriction that no more than 2 training cards appear consecutively. Half of the *S*s were randomly assigned to each order. A 28 × 28 × 45 cm. box, divided into 3 compartments, the middle of which was painted gray, a lateral one blue, and the third green, was used to test *S*s.

The procedure was almost identical to that used by Modigliani (1971). The only difference was that while Modigliani presented the task as one in which *S*s had to tell whether cards were taken from boxes of different colors, the present task was one of putting cards *into* such boxes. In outline, *S* was told in the training phase that each card would go into either the green or the blue compartment of the box in front of him. He was shown an example consisting of a pair of stimuli differing only along the relevant dimension, and was told which of these 2 stimuli went into the green and which into the blue compartment. After *S* examined the stimuli for as long as he wished, he was run to a criterion of 31 out of 32 correct classifications of the training stimuli, one at a time. He was then told that more cards would be shown and that he could now put any of them into the gray compartment if he so wished.

Results and discussion. A placement of a test card into the appropriately colored (blue or green) compartment was defined as conserving and scored as 0. A placement in the gray box was a nonconserving response and was scored as 1. All *S*s but 2 classified all old cards in the conservation phase correctly. In each of the 2 exceptions, *S* made one error that consisted in each case of placing a green card (different for each *S*) in the blue compartment. These errors were considered as temporary slips of memory, and these *S*s' responses to the actual test cards were included in the analysis of the data.

An overall analysis of variance was performed on the data. Only results pertinent to the conceptual boundedness hypothesis will be reported. There were no age effects related to the hypothesis being tested. The mean percentages of conserving responses as a function of perceptual boundedness and transformations in the *other* and the *same* systems, are shown in Table 1. Planned comparisons showed that the 2 mean percentages in the top row did not differ from one another ($F < 1$) but that those in the bottom row did, $F(1, 714) = 19.86$, $p < .01$. Furthermore, the 2 means in the left column did not differ from one another ($F < 1$), while those in the right column did, $F(1, 42) = 21.33$, $p < .01$. Thus, the bottom right mean is significantly different from the other 3 means, which do not differ from each other.

The above results entirely confirmed the conceptual boundedness hypothesis. It seems that the most important aspect of this confirmation is the demonstration that the internal structure of a stimulus, considered as a conceptual system, may affect the way *S*s abstract a simple concept from the larger conceptual framework. Stimulus structure, with the exception of salience (e.g., Trabasso & Bower, 1968) has received little attention in the experimental literature on conceptual behavior. By and large, stimuli are considered as a collection of attributes, some being more salient than others, but having no structural organization of a more complex sort. One exception to this statement is provided by Sutherland (1968). His views were later summarized as follows: "a picture of a man can be decomposed into head, trunk, arms, etc.; head into cheeks, eyes, ears; and at the lowest level, objects are decomposed into segments (or edges) and textures [Sutherland & Mackintosh, 1971, p. 53]." This hierarchical description is remarkably similar to one independently favored by the author with regard to the stimuli used in this experiment. Here the perceptually bounded stimuli are thought of as con-

sisting of (at least) 3 levels. The top level represents the entire stimulus, a face smoking a pipe. The middle level represents the 2 independent systems, the face and the pipe. The bottom level represents the single attributes, the shape of the eyebrows, the brightness of the pipe, etc. Transformations in the present study operated on elements at this bottom level. However, their effect depended on the locus of the attribute(s) operated upon, with respect to the next higher level in the hierarchy, as was the case with perceptually bounded stimuli. While the present results demonstrate in only 1 instance that the abstraction process depends on a level above the one at which the element to be abstracted is located, these results *do* raise the problem of stimulus structure as a potentially important one in the experimental analysis of conceptual behavior.

REFERENCES

Modigliani, V. On the conservation of simple concepts: Generality of the affirmation rule. *Journal of Experimental Psychology*, 1971, **87**, 234–240.

Sutherland, N. S. Outlines of a theory of pattern recognition in animals and man. *Proceedings of the Royal Society, Series B*, 1968, **171**, 297–317.

Sutherland, N. S., & Mackintosh, N. J. *Mechanisms of animal discrimination learning*. New York: Academic Press, 1971.

Trabasso, T., & Bower, G. H. *Attention in learning: Theory and research*. New York: Wiley, 1968.

(Received March 20, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 343–346

SCANNING FOR SIMILAR AND DIFFERENT MATERIAL IN SHORT- AND LONG-TERM MEMORY[1]

C. JAMES SCHEIRER[2] AND MICHAEL J. HANLEY

State University of New York at Binghamton

The *S*s were required to retain a short list of items for an entire experiment while processing successive different short lists of items in the Sternberg paradigm. It was found that reaction time was linearly related to set size for both the permanent list and the immediate lists, but that the slope of the function depended on whether the type of material retained permanently was similar to or different from the successively processed lists. The results are discussed in terms of Sternberg's exhaustive-scanning hypothesis.

The scanning paradigm has been used by Sternberg (1969) and others to investigate search and retrieval processes in memory. In the scanning paradigm, *S* is presented, serially, with a short list of items. Immediately after the presentation of the list, *S* is presented with a single item (probe) and is required to state whether the probe was a member of the list just presented. In this task, performance is practically error free, and response latency is the dependent variable of major interest.

Logically, one would expect the items presented in a scanning task to reside in short-term store (cf. Atkinson & Shiffrin, 1968). Sternberg (1969, Experiment V) and Forrin and Morin (1969) have investigated retrieval of materials in short-term store (STS) and long-term store (LTS) using the scanning paradigm. Sternberg's results led him to postulate that material residing in LTS at the time of the test is transferred item by item to STS, where it is scanned exhaustively. Forrin and Morin, on the other hand, found evidence for independent access to STS and LTS. It must be noted that, although an STS–LTS distinction is discussed by many writers, there does not appear to be an empirical distinction between processing in LTS and processing in STS in the scanning paradigm. Forrin and Morin presented materials assumed to reside in LTS and STS; they found that *S*s were able to access the sets of materials independently. Kaminsky and DeRosa (1972) found that when *S*s were presented a mixed 6-item list of letters and numbers they were able to partition this list, and they appeared to operate on each sublist independently. Thus, for these studies, the utility of a distinction between STS and LTS on other than operational grounds is unclear. Glanzer (1971) has made much the same point in his distinction between LTS and STS as theoretical constructs and long- and short-term memory (LTM and STM) as operational distinctions.

The question thus arises as to the possibility of empirically separating scanning functions in LTS and STS. The present study attempted such a differentiation. It was felt that a critical factor in producing the independent processing found by other *E*s might be the discriminability of the stimulus sets used. Therefore, in this study, the discriminability of stimulus sets was varied systematically.

[1] This research was supported by Grant MH-23231-01 from the National Institute of Mental Health, and by Grant 040-7228A from the Research Foundation of the State of New York to the first author.

[2] Requests for reprints should be sent to C. James Scheirer, Department of Psychology, State University of New York at Binghamton, Binghamton, New York 13901.

Method

The materials used in this study were 2-digit numbers, with each digit chosen randomly from the pool 2–9, and letter bigrams, with each letter chosen randomly from the pool P–W. In no case were both items of a pair identical. Two complete sets of materials were constructed, and half of the *S*s received each set. Items were presented on pairs of IEE display units controlled by an Iconix system controller for .8 sec., with .2 sec. between items. After all of the items on a list had been presented, there was a 2-sec. pause until the test item (probe) appeared. One second prior to the presentation of the probe the word *test*, written in red, appeared above the display units. The probe terminated after 3 sec. There followed an intertrial interval of 11 sec., at which time the word *ready* appeared above the display units. One second later the presentation of the next list of items began. One hundred and twenty such lists were presented to each *S*. The *S* responded to the test stimulus (probe) by depressing 1 of 2 buttons located on a box in front of him. The buttons were labeled *yes* and *no*, corresponding to positive and negative probes.

Experimental conditions. There were 3 experimental conditions, control (C), same (S), and different (D). Condition C was a replication of the basic Sternberg (1969) paradigm. The *S*s received lists of 1, 2, or 3 items, followed 2 sec. later by a probe. An *S* saw only a single type of material. Sixteen *S*s were run in this condition; for 8 *S*s, all lists were composed of digits, while for the other 8 *S*s, all lists were composed of letters. This condition can be conceptualized as a 0 item in the LTM set condition.

Condition S *S*s were required to memorize a list of 1, 2, or 3 items prior to the start of the experiment; this list constituted the LTM set. During the course of the experiment, Condition S *S*s received lists of 1, 2, or 3 items of the *same* type of material they had previously memorized in the LTM set; these items constituted the STM set. Half of the *S*s received lists of numbers in both the STM and LTM sets, and half of the *S*s received lists of letters in both the STM and LTM sets. In Condition *S*, a positive probe was defined by the instructions to include items in either the LTM or STM set; that is, items that had been memorized prior to the start of the experiment and items on the list most recently presented were to be responded to when they were presented as probes by pressing the *yes* button. Negative probes were defined as items not presented in either the LTM or the STM set. Negative probes were, however, always of the same material as that in the LTM and STM sets. Twenty-four *S*s were run in Condition S, 8 with each of the 3 different length LTM lists. The procedures for *S*s in Condition D were identical to those for *S*s in Condition S except that the STM set lists were composed of material of the *opposite* type from that memorized in the LTM set. That is, half of the *S*s received lists of numbers in the LTM set and lists of letters in the STM set, while the other half received lists of letters in the LTM set and lists of numbers in the STM set. Long- and short-term memory items were tested equally often.

For *S*s in Conditions S and D, the LTM set was written in block capital letters on a single 5 [illegible] index card. The *S*s were instructed to memorize the items on the card and were allowed to keep the card with them during the course of the experiment. The *S*s were instructed to respond to probes as rapidly as possible, consistent with errorless performance. There were 64 *S*s, all introductory psychology students at the State University of New York at Binghamton who participated in the experiment as part of the course requirement.

Results and Discussion

The data from the first 12 trials for each *S* were discarded. Mean and harmonic mean latencies for each condition were then computed on the remaining 108 trials for each *S*. Only data from correct responses were utilized; error rates were uniformly low, ranging from 3.6% in Condition C to 2.6% in Condition S and 2.2% in Condition D. Identical patterns of results were found for mean and harmonic mean latencies, so harmonic mean latencies were used throughout this report. These measures were preferred, since they weight short latencies disproportionately. This was felt to be desirable, as latency scores tend to be positively skewed. Harmonic means tend to reduce the size of scores relative to arithmetic means. In general, they differ [illegible] from arithmetic means as the size of the mean decreases. Thus, while differences were small, harmonic mean latencies tended to yield smaller values of reaction time (RT) and slopes steeper than arithmetic means. For example, the best fitting (least squares) straight line relating reaction time to set size (s) for Condition C was RT = 637 + 160s for harmonic mean latencies, and RT = 668 + 172s, for arithmetic mean latencies.

Condition C. The results of this condition, based on the responses of 16 *S*s, replicated the basic Sternberg (1969) results. As is usual in that paradigm, RT was a linear function of list length, with the linear component of the trend accounting for 99.[illegible]% of the variability in the case of positive probes and 99.6% of the variability in the case of negative probes. The best fitting (least squares) straight line was RT = 588 + 158s for positive probes, and RT = 686 + 161s for negative probes. It should be noted that these slopes are considerably larger than those found by Sternberg with single-digit numbers or by Swanson, Johnsen, and Briggs (1972) with pairs of digits. They are also substantially larger than those found previously by this *E* using similar procedures. No reason for the differences obtained in this study are apparent.

Condition D. An overall analysis of variance of the harmonic mean latencies in Condition D showed that positive probes were responded to faster than negative probes, 1029 msec. vs. 1148 msec., $F(1, 12) = 59.39$, $p < .01$; and that list length in the STM set was a significant source of variance, $F(2, 24)$

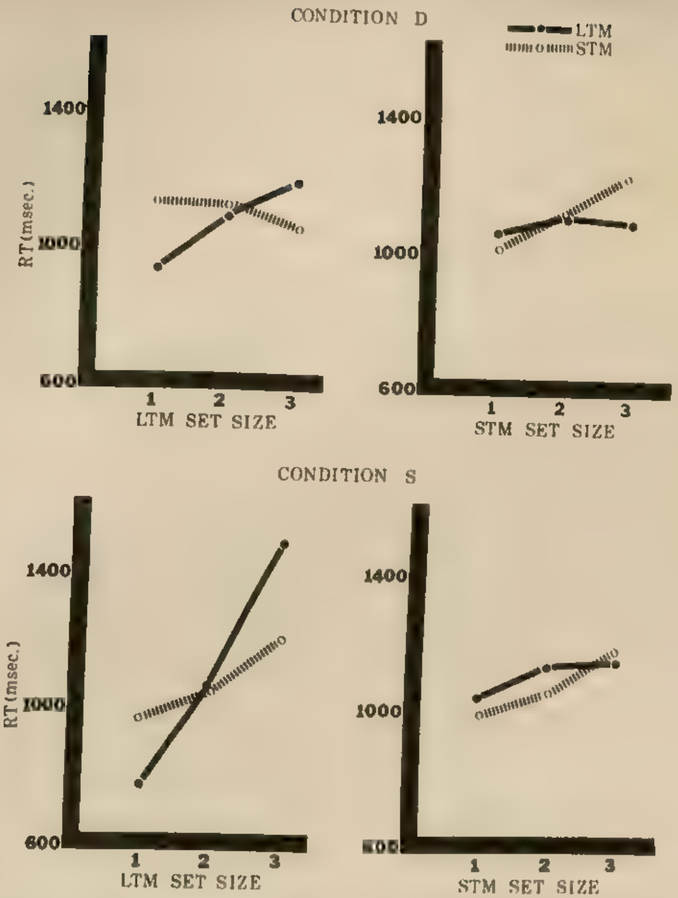


FIGURE 1. Reaction time (RT) for correct responses from the long- and short-term memory (LTM and STM) sets in Conditions different (D) and same (S) as a function of set size.

$= 68.11$, $p < .01$, with 97.2% of the variability accounted for by the linear trend. There were no significant deviations from linearity, F (1, 24) $= 3.74$, $p > .05$. It might be noted that list length in LTS was also a substantial source of variability, but since this factor was between *Ss*, the design lacked power in the test of this component, F (2, 12) < 1. The only other sources of variability to reach significance were the LTM Set List Length × Locus of Test (LTM or STM) interaction, F (2, 12) $= 16.90$, $p < .01$, and the STM Set List Length × Locus of Test interaction, F (2, 24) $= 26.90$, $p < .01$. These interactions, illustrated graphically in the top portion of Figure 1, indicate that memory scanning occurred for both the LTM and STM items, but that changing the size of the memory set in the nontested locus had no effect on speed of recall. The slope of the scanning function in the LTM set was 124 msec/item, while that in the STM set was 104 msec/item. A specific planned comparison, using a combined within-*Ss* error term, indicated that these slopes did not differ reliably, F (1, 132) $= 1.86$, $p > .05$; and that the size of the memory load in the nontested store did not affect the speed of recall, F (4, 132) < 1.

The data presented in Figure 1 suggest that *Ss* access the STM and LTM material independently when that material is highly discriminable. This conclusion follows from the fact that in Condition D there is no effect of LTM set size on scanning in the STM set, and no effect of STM set size on scanning in the LTM set. These results thus replicate the results of Forrin and Morin (1969). It should be noted, however, that although it would seem likely that STM and LTM sets would reside in different memory states (STS and LTS) because of the amount of information and operations involved in the task, this conclusion is not forced by the data. It is possible that all items are held in active memory, but that discriminable subsets of items are acted on independently (Kaminsky & DeRosa, 1972).

Condition S. An overall analysis of variance on harmonic mean latencies revealed that list length in the LTM set was a significant source of variance,

F (2, 12) = 5.53, $p < .05$, with 97.0% of the variability in the effect accounted for by the linearity of the trend. Deviations from linearity failed to reach significance, F (1, 12) < 1. Harmonic mean latencies of the 3 types of test items also differed significantly, F (2, 24) = 21.82, $p < .01$, with means of 1071 msec. for positive probes in the STM set, 1102 msec. for positive probes in the LTM set, and 1243 msec. for negative probes. Specific comparisons among these conditions revealed that negative probes were responded to reliably slower than positive probes, F (1, 24) = 42.15, $p < .01$, but RT to positive probes did not differ according to memory set, F (1, 24) = 1.24, $p > .05$. List length in the STM set was also a significant source of variability, F (2, 24) = 20.35, $p < .01$, with the linear component of trend accounting for 99.98% of the variability in the effect. The only other sources of variability to reach significance were the LTM Set List Length × Test Conditions interaction, F (4, 24) = 15.83, $p < .01$, and the STM Set List Length × Test Conditions interaction, F (4, 48) = 3.14, $p < .05$. These effects are presented graphically in the bottom portion of Figure 1.

The function relating RT to set size in the LTM set in Condition S was RT = 380 + 361s. This function was basically linear in form, accounting for 99.3% of the variability associated with difference in set size. The function obtained for this condition has a slope substantially larger and an intercept substantially smaller than that found in any other condition. Using Sternberg's (1969) interpretation of slope and intercept values, the LTM items in Condition S are accessed faster than those in any other condition (including Condition C, access to a single set of items), but are scanned more slowly. While the slow scanning rate for LTM items appears plausible, and in fact replicates Sternberg (Experiment V), the very fast access time appears difficult to handle within Sternberg's model.

One possible explanation of this data is that *Ss* might be disregarding the classification of items into LTM and STM sets. The best fitting straight line through these points is RT = 657 + 104s, which accounts for 91.2% of the variability in prediction. One interpretation of this function is that material from the LTM set is held in the same locus as items in the STM set; the total set is then scanned exhaustively before a decision is made concerning the STM set items. The material from the LTM set could not be the actual LTM items, however, since this would predict identical scanning functions for the LTM and STM sets.

An alternative explanation of the data from Condition S suggests that *Ss* were able to differentiate between the LTM and STM set only when there was a single item in the LTM set. This interpretation is suggested by the data of Darley, Klatzky, and Atkinson (1972), which indicate that *Ss* are able to process items which have a high probability of being tested, independent of the remaining items on a list. In this study, when there was a single item in the LTM set, it was tested 25% of the time (50% positive probes, 50% LTM tests) a proportion of testing that might lead to this item being treated as a high-probability test item. This explanation is not supported, however, by the fact that a similar facilitation did not occur in Condition D.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. New York: Academic Press, 1968.

Darley, C. F., Klatzky, R. L., & Atkinson, R. C. Effects of memory load on reaction time. *Journal of Experimental Psychology*, 1972, **96**, 232–234.

Forrin, B., & Morin, R. E. Recognition times for items in short and long-term memory. *Acta Psychologica*, 1969, **30**, 126–141.

Glanzer, M. Storage mechanisms in recall. In G. H. Bower (Ed.), *The psychology of learning and motivation.* Vol. 5. New York: Academic Press, 1971.

Kaminsky, C. A., & DeRosa, D. V. Influences of rehearsal cues and set organization on short-term recognition memory. *Journal of Experimental Psychology*, 1972, **96**, 449–454.

Swanson, J. M., Johnsen, A. M., & Briggs, G. E. Recoding in a memory search task. *Journal of Experimental Psychology*, 1972, **93**, 1–9.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

(Received March 26, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 347-349

INFLUENCE OF ACTIVE AND PASSIVE VOCALIZATION ON SHORT-TERM RECALL[1]

PHILLIP M. TELL[2] AND ALEXANDER M. FERGUSON

Florida Technological University

The Brown–Peterson distractor technique was used to investigate the effects of active and passive vocalization vs. silent visual presentation on recall at various retention intervals. Under active vocalization conditions, visually presented stimulus items were voiced aloud by *S*, and under passive vocalization conditions, these items were aurally presented by *E*. The findings indicated that both active and passive vocalization produced better recall than the silent presentation condition at retention intervals of up to 5 sec. Also, the active and passive vocalization conditions differed mainly at the longer retention intervals, with recall much higher under the passive vocalization condition. There were only minimal differences between active and passive vocalization at the short retention intervals.

There is currently much theoretical interest in the effects of various auditory and articulatory stimulus presentation conditions on short-term recall. In order to facilitate understanding, the present study will maintain Crowder's (1970) distinction between active and passive vocalization presentation conditions. Active vocalization refers to presentation conditions where the to-be-remembered stimulus items are voiced aloud by *S* as they are visually presented. Under passive vocalization conditions, *S* listens to *E* read each stimulus item as it is visually presented.

The findings of many experiments indicate that both auditory presentation of stimulus items by *E* (passive vocalization) and vocalization of these items by *S* (active vocalization) have significant advantages over silently read, visually presented items (Cooley & McNulty, 1967; Crowder, 1970; Murdock, 1966, 1967; Murray, 1965; Tell, 1971). This advantage of active vocalization over silent visual presentation has been demonstrated both with the Brown–Peterson distractor task (Brown, 1958; Peterson & Peterson, 1959) and the immediate memory task. It seems reasonable to assume that the same memory systems can be investigated through the use of the distractor task or the immediate memory task. Thus, recall performance at short retention intervals (distractor task) is assumed to reflect the same memory characteristics as recall at recency positions of the serial position curve. Likewise, it is assumed that recall after longer retention intervals should be comparable to the primacy positions of the serial position curve. When measured with the distractor task, active vocalization is superior to silent visual presentation mainly at retention intervals of less than 7 sec. (Tell & Voss, 1970). When measured with an immediate memory task, active vocalization is generally superior to silent visual presentation on the last few serial positions only (Conrad & Hull, 1968). The same general types of results also occur when passive vocalization is compared with silent visual presentation (Corballis, 1966; Crowder, 1970; Grant & McCormack, 1970). Thus, it strongly appears that the same information system can be investigated with both the distractor task and the immediate memory task.

Currently, the only available direct comparison of the effects of active and passive vocalization on immediate recall is by Crowder (1970). His most relevant finding in terms of the present discussion was that passive vocalization was superior to active vocalization for only the first few serial positions (primacy positions). Both of these conditions were superior to the silent visual presentation condition at the recency positions only.

It has previously been speculated that the auditory feedback that results from active vocalization is functionally equivalent as to the auditory input resulting from passive vocalization (Tell & Voss, 1970) as a source of information. Crowder's (1970) data indicate that this is clearly not the case. As Crowder speculates, active vocalization during presentation could interfere with effective rehearsal or encoding strategies compared to a passive condition, in which *S* is free to engage in whatever type of encoding activity is most effective. This may account for the apparent superiority of passive vocalization at the primacy positions.

The present study was designed to provide a finer analysis of the effects of active and passive vocalization vs. silent visual presentation on recall at various retention intervals through the use of the distractor technique. If recall performance at short retention intervals indeed reflects the operation of the same memory characteristics as does recall of recency items, then it would be predicted that both active and passive vocalization should be superior to silent visual presentation at these short retention intervals. Likewise, it is predicted that passive vocalization will be superior to active vocalization at longer retention intervals, which correspond informationally to the primacy positions of the serial position curve.

[1] This research was supported by National Institute of Mental Health Grant MH-22282 to the first author. The authors are very grateful to Robert Crowder for comments on an earlier version of this paper.

[2] Requests for reprints should be sent to Phillip M. Tell, Department of Psychology, Florida Technological University, Box 25000, Orlando, Florida 32816.

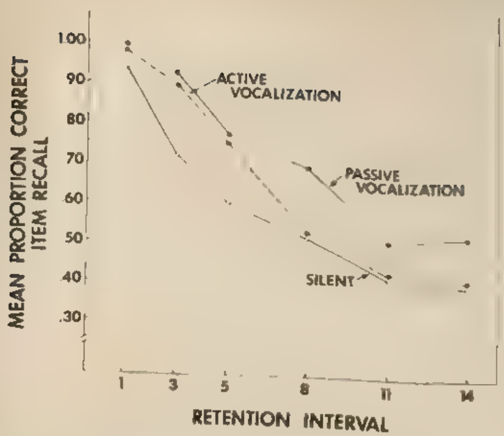


FIGURE 1. Mean proportion correct item recall as a function of presentation condition and retention interval.

Method. The *S*s were 90 undergraduates at Florida Technological University. All *S*s were tested individually and received course credit for participating in the experiment.

Twenty-eight consonant (C) trigrams were selected from the Witmer norms in Appendix B of Underwood and Schulz (1960) in such a way as to minimize association value and acoustic similarity. Repeated occurrences of any consonant were separated by at least 2 CCCs. Four of the 28 CCCs served as practice items and were not used in analysis of data or counterbalancing of conditions. The CCCs were photographed on 35-mm. film and mounted in slide holders. Blue gelatin filters mounted in slide holders marked the onset of each trial, and a series of 3 question marks served as the recall signals.

The slides were projected by means of a Kodak Carousel 800 projector. Series of 2-digit numbers for the interpolated task were filmed on 16-mm. motion picture film and projected with a Dunning Animatic film strip projector.

Both of these projectors were operated by tone pulses recorded on magnetic tape by the use of electromechanical programming equipment. One channel of a Sony Model 630 stereo tape recorder was used for this purpose. A Kodak Carousel programmer and sound synchronizer were used for both the recording and the playback of tones.

The second channel contained *E*'s recorded vocalizations of the CCCs, calibrated to coincide approximately with the visual display of the CCCs. This channel was turned on in the passive vocalization condition and was turned off under the silent and active vocalization conditions.

Using the Brown–Peterson distractor technique (Brown, 1958; Peterson & Peterson, 1959), *S*s attempted to retain CCCs while performing an interpolated task. The interpolated task involved the presentation of digit pairs at a rate of 1 sec/pair, and all *S*s were tested after 1, 3, 5, 8, 11, and 14 pairs of digits. The interpolated activity required *S* to verbally classify each member of the digi[t] as odd or even, from left to right.

Each *S* received 28 trials: the first 4 were [prac]tice and the rest were test trials. The o[rder of] trigram presentation was the same for all *S*s. [This] confounds practice completely with individua[l tri]grams, but balances practice against rete[ntion] interval.

The 90 *S*s were divided into 3 presentation [con]ditions. In the silent presentation condition, [ma]terials were presented visually and *S*s were request[ed] to silently read the trigrams to themselves. The active vocalization condition required *S*s to read the trigrams aloud while they were on the screen. Under the passive vocalization conditions, *S*s heard *E* read the trigrams aloud while they were presented on the screen. Thus, during the 24 scored trials in each of the 3 presentation conditions, a given *S* received 4 test trials at each of the 6 retention intervals.

A 6 × 6 Latin square was used to order the 6 retention intervals within the 24 test trials. [Each *S*] received an ordering of retention intervals d[eter]mined by 4 rows of the Latin square. Th[ree] separate samples composed of 4 rows each w[ere] used, so that each retention interval followed [all] other retention intervals an equal number of t[imes] and never followed itself. Within any presentation condition, 10 different *S*s received each of the 3 retention interval orders.

Briefly, a single trial proceeded in the following manner. A blue ready light illuminated the screen for 1 sec. Following this by .75 sec., a CCC was presented on the screen for .75 sec., during which time *S* either read it silently, spoke it aloud, or listened to it read by *E* over a speaker. After the presentation of each CCC, *S*s were tested at one of the 6 retention intervals (1, 3, 5, 8, 11 or 14 sec.). The interpolated activity was continually scored, and when necessary, *S*s were reminded to do as well as possible. Three question marks appeared after the last digit pair was removed from the screen. These remained on the screen for t[he] 10 sec. during which *S* attempted to recall the C[CC]. A rest interval of 10 sec. separated the trials.

Results and discussion. The data of each *S* were scored in terms of the number of consonants correctly recalled, regardless of whether they appeared in the correct intratrigram position. With low-association-value trigrams, it appears that the item score is the most meaningful and sensitive unit of analysis (Wickelgren, 1965). Results are not presented for ordered scores because they are in close agreement with item scores in all important respects.

An overall analysis of variance was performed on the recall scores. This analysis indicated that the 3 types of presentation conditions resulted in significantly different recall functions, F (2, 87) = 4.65, $p < .05$. As can be seen in Figure 1, recall performance for all groups decreased as a function of retention interval, F (5, 435) = 316.4, $p < .001$. It is apparent that the types of presentation conditions had different effects at the various retention intervals tested, F (10, 435) = 4.0,

$p < .01$. Both active and passive vocalization conditions are clearly superior to silent visual presentation at retention intervals of up to 5 sec. This might be assumed to result from the additional echoic information provided by the auditory input from these presentation conditions (Neisser, 1967). However, the initial differences between active vocalization and silent visual presentation disappear at the longer retention intervals (8, 11, and 14 sec.). It is also apparent that the passive vocalization condition provides higher recall than either of the other 2 presentation conditions at these longer retention intervals.

The present findings closely parallel those obtained with the immediate memory task (Crowder, 1970). First, both active and passive vocalization conditions produced better recall than the silent presentation condition at retention intervals of up to 5 sec. Second, the active and passive vocalization conditions differed mainly at the longer retention intervals. There are just minimal differences due to active or passive vocalization at the short retention intervals. The evidence strongly indicates that the same memory systems can be investigated using either supraspan memory units (immediate memory task) or subspan memory units (distractor task).

It appears that the requirement of active vocalization at presentation interferes with effective encoding operations. Active vocalization may demand more attention at presentation than does passive vocalization. The effect may not involve just the disruption of rehearsal strategies, but may instead lessen the selective attention capabilities that are necessary for effective encoding in memory. Thus, the advantages of passive vocalization are mainly at the longer retention intervals, when the information in echoic memory has dissipated. The pr[illegible]a clearly indicate that active and [illegible]ization conditions are not comparable [illegible]nput to the memory system when c[illegible]with silent visual presentation.

REFERENCES

Brown, J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

Conrad, R., & Hull, A. J. Input modality and the serial position curve in short-term memory. *Psychonomic Science*, 1968, **10**, 135–136.

Cooley, R. K., & McNulty, J. A. Recall of individual CCC trigrams over short intervals of time as a function of mode of presentation. *Psychonomic Science*, 1967, **9**, 543–544.

Corballis, M. C. Rehearsal and decay in immediate recall of visually and aurally presented items. *Canadian Journal of Psychology*, 1966, **20**, 43–51.

Crowder, R. G. The role of one's own voice in immediate memory. *Journal of Cognitive Psychology*, 1970, **1**, 157–178.

Grant, K. W., & McCormack, P. D. Auditory and visual short-term memory with successive presentation both modalities. *Psychonomic Science*, 1970, **17**, 341–342.

Murdock, B. B. Visual and auditory stores in short-term memory. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 206–211.

Murdock, B. B. Auditory and visual stores in short-term memory. *Acta Psychologica*, 1967, **27**, 316–324.

Murray, D. J. Vocalization at presentation and immediate recall with varying recall methods. *Quarterly Journal of Experimental Psychology*, 1965, **17**, 47–56.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

Tell, P. M. Influence of vocalization on short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 149–156.

Tell, P. M., & Voss, B. X. The influence of vocalization activity on short-term recall. *Psychonomic Science*, 1970, **21**, 120–121.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

Wickelgren, W. A. Acoustical similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 1965, **70**, 102–108.

(Received March 31, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 2, 350–352

BELOW-ZERO CONDITIONED INHIBITION OF THE RABBIT'S NICTITATING MEMBRANE RESPONSE[1]

HORACE G. MARCHANT, III, AND JOHN W. MOORE[2]

University of Massachusetts

Albino rabbits received Pavlovian conditioned inhibition training of their nictitating membrane response to a light (positive conditioned stimulus, CS+) and the same light plus a pure tone (negative conditioned stimulus, CS−). Subsequent acquisition to the tone was retarded relative to naive and "sit" controls as well as groups receiving differential conditioning to noise (CS+) and light plus tone (CS−), and light (CS+) and light plus noise (CS−). The results were interpreted as demonstrating that conditioned inhibition training results in a CS with a negative (below-zero) value.

Although interest in Pavlovian conditioned inhibition has increased in recent years, there remain problems of interpretation of this phenomenon (Hearst, 1972). Conditioned inhibition is usually defined as a response tendency opposite that of excitation, where excitation is the increase in the strength of a conditioned response (CR) due to pairings of the conditioned stimulus (CS) and unconditioned stimulus (UCS). This definition of excitation implies that conditioned inhibition is a tendency to not respond resulting from presentation of the CS without reinforcement. The question of interpretation is whether failure to respond in this case may not be more parsimoniously conceptualized as resulting from a loss of excitation rather than from an active inhibitory process (cf. Brown & Jenkins, 1967).

The assumption that both excitatory and inhibitory forms of conditioning exist, separated by a point of zero-response strength, is a crucial feature of a number of learning theories, including familiar Hullian theories of discrimination learning (Spence, 1937). The theory of classical conditioning developed by Rescorla and Wagner also assumes that CR strength (V in their model) may take on either positive or negative value (cf. Wagner & Rescorla, 1972). The demonstration that conditioned inhibition results in a CS of negative value would seem to be of some importance for both of these theories.

The following 3 criteria should be satisfied in order to assert that a given training experience results in a CS with inhibitory properties: (*a*) a zero point must be established, (*b*) CR strength must lie below this point, and (*c*) this below-zero or negative effect must be the result of learning rather than some nonassociative factor such as an attention decrement. Marchant, Mis, and Moore (1972) investigated conditioned inhibition of the rabbit's nictitating membrane response (NMR) using Pavlov's procedure of reinforcing one stimulus (CS_1) while systematically nonreinforcing a compound consisting of CS_1 and another stimulus (CS_2). Although successful in demonstrating inhibitory summation and retardation (cf. Rescorla, 1969), this experiment failed to establish conditioned inhibition as a below-zero effect. One reason for this was that all *S*s were conditioned to CS_2, the potential inhibitor, prior to differential conditioning in an attempt to circumvent the development of latent inhibition to CS_2. Therefore, retardation was assessed in terms of rate of reacquisition following training rather than rate of initial acquisition. Secondly, control *S*s received prior conditioning to CS_2 followed by pseudoconditioned inhibition training in which each of the 2 trial types was paired with the UCS 50% of the time in a random sequence. Under these circumstances it seems unlikely that CS_2 in the control group defined a zero-reference point. Wagner and his associates have also investigated conditioned inhibition of the rabbit's eyeblink. These studies, while supporting the Rescorla and Wagner model, did not address the question whether conditioned inhibition is a below-zero effect (cf. Wagner & Rescorla, 1972).

The present experiment sought to establish conditioned inhibition of the rabbit's NMR as a below-zero effect by utilizing essentially the same training procedures employed by Marchant et al. (1972) but by comparing initial acquisition of the conditioned inhibitor with a reference group not subject to the complications mentioned above. Specifically, conditioning to CS_2 following inhibitory training was contrasted with a group which merely sat restrained in the conditioning chambers during the conditioned inhibition phase. Additional controls, not employed previously, were included to assess the role of a variety of potential artifacts contributing to retardation of acquisition to the inhibitory CS.

Method. The *S*s were 49 experimentally naive rabbits. Up to 4 *S*s were run simultaneously in ventilated soundproofed file drawers, and the details of recording the conditioned NMR have been described elsewhere (Marchant et al., 1972). The CSs were a 1,200-Hz. pure tone (T) of 86 db. SPL, 2 4.5-v. incandescent lights (L) behind milk

[1] This research was supported by National Science Foundation Grant GB-24557. This experiment was based on the first author's master's thesis. We wish to thank John J. B. Ayres and D. Anderson for their contributions to the design of this study.

[2] Requests for reprints should be sent to John W. Moore, Department of Psychology, University of Massachusetts, Middlesex House, Amherst, Massachusetts 01002.

TABLE 1

Summary of Experimental Design

Group	Stage 1	Stage 2
Naive ($N = 7$)	—	T+
Sit ($N = 14$)	sit	T+
L+LT− ($N = 14$)	L+ LT−	T+
L+LN− ($N = 7$)	L+ LN−	T+
N+LT− ($N = 7$)	N+ LT−	T+

Note. Abbreviations: T = tone, L = light, N = noise burst, + = reinforced, − = nonreinforced.

white screens located in front of *S*, and an 86 db. SPL burst of white noise (N). The duration of each CS was a constant 550 msec. The UCS was a 2-ma. ac shock of 50-msec. duration delivered through stainless steel wound clips attached to the infraorbital region of the right eye. The intertrial interval was a constant 30 sec. and the interstimulus interval was 500 msec. throughout.

The experimental design is summarized in Table 1. One day prior to Stage 1 training (or, in the case of Group Naive, Stage 2 training), *S*s had their right NMs sutured and were habituated to restraint and the apparatus for a period of approximately 15 min. For *S*s in Group L + LT−, Stage 1 training consisted of 50 reinforced presentations of the L stimulus, unsystematically intermingled with 50 nonreinforced presentations of the simultaneous compound L plus T, in each daily session. This training was expected to produce an inhibitory T stimulus and was continued until the following criteria had been met: Between Sessions 6 and 12, if (*a*) percentage of CRs to CS− had first increased and then decreased over sessions and (*b*) percentage of CRs to CS+ was better than 80% and the percentage of CRs to CS− was less than 30%, then Stage 2 training began for that *S* midway through the next day's session. Group Naive did not participate in the experiment until Stage 2.

Group Sit served as a control for any possible effects of handling, suturing, sitting in restraint, experience with static apparatus cues, etc. In Stage 1, *S*s in this group were yoked to *S*s in Group L + LT− and were treated in an identical manner but without CS and UCS presentations. Group L + LN− was trained to the same criterion as Group L + LT−, but the discrimination was between L+ and (L + N)− as opposed to L+ and (L + T)− as in the Group L + LT−. Group L + LN− served as a control for experience with CSs, the UCS, conditioning and, more specifically, inhibitory training (i.e., inhibitory training per se may be sufficient to produce a deficit in acquisition of all stimuli). Group N + LT− was also run to the same criterion as Group L + LN−, and served as a control for experience with the LT− compound and generalization decrement in Stage 2.

The same nonsystematic order of CS+ and CS− trials in Stage 1 was used for Groups L + LT−, N + LT−, and L + LN−.

Stage 2 training for all groups began with 50 reinforced trials to T in the last half of the last session of Stage 1. This was followed by 100 trials 5 (±½) hr. later and 2 100-trial sessions the following day at approximately the same times as the previous day's sessions.

Results. All *S*s in Groups L + LT−, N + LT−, and L + LT− met the discrimination criteria in Stage 1 within the allotted 12 sessions. Acquisition to T in Stage 2 was assessed in terms of (*a*) trials to a criterion of 80% CRs over 3 consecutive blocks of 10 trials and (*b*) percentage of CRs over the initial 150 acquisition trials. Table 2 shows the mean of each measure of Stage 2 conditioning for each group.

Analysis of variance indicated significant ($p < .001$) differences among groups for both measures with $F(4, 44) = 19.56$, and $F(4, 44) = 13.91$, for trials to criterion and percentage of CRs, respectively.

Individual planned t tests based on the analyses of variance ($df = 44$) indicated that Group L + LT− was significantly slower to reach criterion than Group Sit, $t = 4.15$, $p < .001$. Group L + LT− was also slower to acquire than Group Naive, $t = 4.07$, $p < .001$. Description of the results of individual pair-wise contrasts among groups is confined to the trials to criterion measure since similar comparisons for percentage of CRs over the first 150 acquisition trials were significant at better than the .025 level.

Groups N+LT− and L+LN−, which did not differ significantly from each other, conditioned to T significantly more quickly than Group Sit, $t = 3.81$, $p < .001$, and $t = 3.94$, $p < .001$, respectively. Groups N+LT− and L+LN− also conditioned to T significantly more quickly than Group Naive, $t = 2.58$, $p < .01$, and $t = 2.70$, $p < .005$, respectively. Similar comparisons for percentage of CRs over the first 150 acquisition trials were also significant at better than the .005 level.

Discussion. The principal findings of this experiment were as follows. (*a*) Pavlovian conditioned inhibition training (L+LT−) resulted in retarded acquisition of the inhibitor (T) when compared with "sit" and naive controls. In this sense, conditioned inhibition produced a CS with negative or below-zero value as required by theo-

TABLE 2

Mean Trials to Criterion and Percentage of Conditioned Responses (CRs) in Stage 2

Group	Trials to criterion	Percentage of CRs
L + LT−	134.4	35.3
Sit	81.6	48.5
Naive	68.6	56.2
L + LN−	21.3	81.8
N + LT−	19.1	82.5

ries (Spence, 1937; Wagner & Rescorla, 1972) which assume the existence of active inhibitory as well as excitatory processes. (*b*) Differential conditioning in an N+LT− paradigm and conditioned inhibition training of a noise burst (L+LN−) both resulted in faster subsequent acquisition to T than in the sit and naive control groups.

The below-zero effect of conditioned inhibition training cannot be readily interpreted as resulting from an attention decrement since the criteria of differential conditioning imposed in Stage 1 of training assured attention to the conditional inhibitor right up to the onset of acquisition in Stage 2. If attention decrement were responsible for retardation of acquisition to T in Group L+LT−, retardation would have been observed in Group N+LT− as well. Failure to observe retardation to T in the latter group cannot be attributed to "overshadowing" of T by L since independent observations in our laboratory have established that T is the more effective CS in simple conditioning and hence the more "salient" of the 2 stimuli.

Repeated preexposure to the CS or UCS is known to retard conditioning of the rabbit's NMR (e.g., Mis & Moore, 1973). Groups N+LT− and L+LN− served as controls for these factors in the present experiment. The fact that conditioning to T in Groups L+LT− was substantially retarded relative to these 2 groups would seem to eliminate preexposure as the mechanism of retardation.

Failure to observe retardation to T in Groups N+LT− is consistent with the prediction of the Rescorla-Wagner model that inhibition develops as a consequence of nonreinforcement against a baseline or background of excitation (Wagner & Rescorla, 1972). In the present experiment, T in Group L+LT− satisfied this stipulation by virtue of the fact that nonreinforcement of T occurred in the presence of an excitatory CS, L, and retardation of acquisition to T was observed in this case. By contrast, L could not provide the excitatory component necessary for the development of inhibition to T in Group N+LT−.

Finally, Groups L+LN− acquired a CR to T more rapidly than any group other than N+LT−. Since both N and T are in the same modality, one might have expected inhibition to transfer from Stage 1 (N) to Stage 2 (T) and result in some retardation. That such transfer did not occur may have been due either to (*a*) excitatory transfer overcoming any inhibitory transfer present or (*b*) overshadowing of the L by the N so that Group L+LN− would be, in effect, a simple differential conditioning group. In the latter case no active inhibition would be produced due to the absence of an excitatory component.

REFERENCES

Brown, P. L., & Jenkins, H. M. Conditioned inhibition and excitation in operant discrimination learning. *Journal of Experimental Psychology*, 1967, **75**, 255–266.

Hearst, E. Some persistent problems in the analysis of conditioned inhibition. In R. S. Boakes & M. S. Halliday (Eds.), *Inhibition and learning*. London: Academic Press, 1972.

Marchant, H. G., III, Mis, F. W., & Moore, J. W. Conditioned inhibition of the rabbit's nictitating membrane response. *Journal of Experimental Psychology*, 1972, **95**, 408–411.

Mis, F. W., & Moore, J. W. Effects of preacquisition UCS exposure on classical conditioning of the rabbit's nictitating membrane response. *Learning and Motivation*, 1973, **4**, 108–114.

Rescorla, A. Pavlovian conditioned inhibition. *Psychological Bulletin*, 1969, **73**, 77–94.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, 1937, **44**, 430–444.

Wagner, A. R., & Rescorla, R. A. Inhibition in Pavlovian conditioning: Applications of a theory. In R. A. Boakes & M. S. Halliday (Eds.), *Inhibition and learning*. London: Academic Press, 1972.

(Received April 2, 1973)

JOURNAL OF EXPERIMENTAL PSYCHOLOGY
MONOGRAPH

VOL. 102, No. 2, 353–376 FEBRUARY 1974

THE APPARENT MAGNITUDE OF NUMBER SCALED BY RANDOM PRODUCTION [1]

WILLIAM P. BANKS [2] AND DAVID K. HILL [3]

Pomona College

A technique for generating a ratio scale for any psychophysical continuum by having *S*s generate subjectively random productions on that continuum is described and used to scale numerical magnitude, duration, and line length. The perceived magnitude of numbers grows as an increasingly decelerated function of their absolute magnitude, but it is approximately logarithmic from 10 to 1,000. The range of 1–500 is fitted well with a power function with an exponent of about $\frac{1}{3}$. Several experiments defend the position that the psychological scale of number is compressive and explore a number of possible artifacts in the random-production technique. Cross-modality matches between numbers, line lengths, and durations are predicted from the random-production scales derived for each of these dimensions. The predictions give power functions whose exponents are in some agreement with previous experimental estimates. Random-production scales for handgrip force, intensity of electric shocks, and subjective area are consistent with their magnitude-estimation scales, given that number has an exponent of $\frac{1}{3}$.

In magnitude-estimation experiments, *S* assigns numbers to stimuli according to some rule, and *E* must take him at his word on the rule he is using. Of course, there is no reason to assume that *S* is lying about his perceptions, but there is no way we can be sure that he has the same conception of the magnitude of the response numbers he is using that *E* (or arithmetic) does. Thus, if *S* calls a given stimulus "20," the magnitude-estimation technique assumes its subjective magnitude is twice that of the stimulus called "10." But *S*'s conception of the numerical magnitude could be such that a stimulus subjectively twice as great as the one called "10" would have to be called "30." In this case, the resulting scale, while replicable and capable of entering properly into cross-modality matches, would not reflect *sensory* magnitudes accurately.

The question of the form of the psychological scale of number is central to understanding the meaning of direct-scaling experiments; yet, while these experiments continue to proliferate and conclusions about sensory function continue to be drawn from them, no agreement on this important question has been reached. Perhaps the simplest and most convenient approach is to assume that the magnitude

[1] Experiments I–V were performed by the first author under a National Institute of Mental Health predoctoral training grant administered by Johns Hopkins University, U.S. Public Health Service Postdoctoral Grant 5T01-GM01207-05 at the Institute of Human Learning, University of California, Berkeley, and with the assistance of a Pomona College faculty research grant. Some of the writing was done with the support of a Pomona College summer research fellowship. Experiments VI–IX formed part of the second author's senior honors thesis at Pomona College and were supported by a Pomona College faculty research grant to the first author. Of the many who have made valuable suggestions on this work, P. Arabie, J. M. Myhre, E. A. C. Thomas, and W. S. Torgerson deserve particular thanks. The astute critical comments of R. D. Luce, L. E. Marks, and J. C. Stevens also contributed greatly to the paper, although the authors must apologize for not following their advice in every case.

[2] Requests for reprints should be sent to William P. Banks, Department of Psychology, Pomona College, Claremont, California 91711.

[3] Now at the University of Oregon Law School.

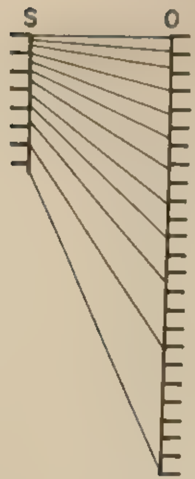


FIGURE 1. Psychophysical mapping between O, the observable continuum, and S, the continuum of subjective experience.

scale of number is linear. There is, however, evidence that it is not (cf. Atteneave, 1962; Curtis, 1970; Ekman & Hosman, 1965; Fairbank, 1969; Moyer & Landauer, 1967; Rule, 1969; Sekuler, Rubin, & Armstrong, 1971), and in any event it would be better to demonstrate its linearity before proceeding on the assumption that it is linear. (For evidence *for* linearity, see Dawson, 1971, and Goude, 1962.)

Previous attempts to scale the psychological number continuum have shown it to be compressive, but they are in some disagreement over the precise form of the function. Some seem to have shown that the scale is logarithmic (Moyer & Landauer, 1967), and others, that it is a power function (Curtis & Fox, 1969). These conflicting results reflect the difficulties intrinsic to the problem. Conventional direct-scaling techniques cannot be used in unmodified form because with them, one dimension must be scaled in terms of another. Whatever function is obtained will be a result of the input and output dimensions working together. Conventional indirect methods cannot be applied in unmodified form to scaling the numbers, either. One obvious difficulty is that there is no just noticeable difference for number (2 numbers are either equal or they are not), and thus there is no discrimination problem to exploit in constructing a scale. (However, for an ingeniously devised Thurstonian scale of number, see Rule, 1969).

This article reports experiments utilizing a new technique for studying how the magnitude of numbers is perceived. The technique can be used to scale any dimension *S* can control, and it is not limited to the numerical continuum. The technique is very simple: *S*s are asked to produce random quantities—in this case, numbers—as rapidly as possible, with no fixed upper limit, in no special order, and as randomly as possible. The assumption behind the random-number-production technique is that *S*s will display, by the form of the distribution they produce, their conception of numerical magnitude. It is as though *S*s refer to their internal "ruler"—presumably the same one that is used in magnitude estimation—and select numbers from this source in a nonsystematic fashion. Their sampling from this line is assumed to be rectilinear and random. By this reasoning, intervals will be represented in proportion to their psychological size. If, for example, the numbers 105 and 115 represent points that are about as far apart psychologically as 5 and 15, *S*s should generate, on the average, as many numbers in one of these intervals as in the other.

The assumption that *S*s sample rectilinearly from their subjective scale in the number-production task permits a powerful analysis of the results. If their sampling is in fact rectilinear, the expected value of the order statistic for the *n* random samples can give a ratio scale metric for the psychological continuum over which the samples were made, as the following considerations show.

In a random-production experiment on any continuum, *S* produces *n* subjective randomly selected quantities on that continuum. If the productions of each *S* are put in order from smallest to largest and an average of the productions at each ordinal position is taken across *S*s, then the resulting *n* averages estimate the expected values of the order statistics for those samples (Hogg, 1969). If sampling

on the continuum is uniform, that is if $f(x) = 1,\ 0 < x < 1$, then the probability density function of Y_k, the kth of the n order statistics on $0 < x < 1$, is

$$g_k(y) = \frac{n!}{(k-1)!(n-k)!} \cdot (y)^{k-1}(1-y)^{n-k}.$$

The expected value of Y_k can thus be calculated as follows:

$$E[Y_k] = \int_0^1 y g_k(y)dy$$

$$= \int_0^1 \frac{n!}{(k-1)!(n-k)!} \cdot y^k(1-y)^{n-k}dy$$

$$= \frac{n!}{(k-1)!(n-k)!} \cdot \frac{\Gamma(k+1)\cdot\Gamma(n-k+1)}{\Gamma(n+2)}$$

$$= \frac{n!}{(k-1)!(n-k)!} \cdot \frac{k!(n-k)!}{(n+1)!}$$

$$= k\frac{1}{n+1}.$$

Thus, the order statistics divide the 0–1 line into n equally spaced points with a unit difference of $1/(n+1)$. That is, the kth order statistic will be k units of size $1/(n+1)$ from the bottom. If the line extended from zero to N, the unit difference would then be $N/(n+1)$.

It will be assumed throughout this article that *Ss* map all continua, in the process of perceiving them, onto a "subjective" scale which, for the sake of convenience, will be represented as a line going from zero to one. Figure 1 illustrates this mapping. In the figure, the O scale represents the physical (observable) continuum, and the S scale represents the subjective mapping of this continuum. The mapping function, M(O), assumed in this figure is monotonic and compressive but, in general, it may not be compressive for all continua. In the case of numbers, "infinity" may be mapped on the subjective continuum; if it is, the subjective continuum, being finite, must become infinitely compressed at the top.

When *Ss* are asked to sample rectilinearly from a continuum, they must do this sampling on the S continuum, for this is all that is available to them. Thus, to the extent that *Ss* can obey the rectilinearity instructions, the order statistics derived from n random samples will determine n points on the S continuum, equally spaced with a unit proportional to $1/(n+1)$. The S values are not, of course, directly available, but the order statistics that result from a random-production experiment provide n values on the O continuum that correspond to n equally spaced values on the S continuum. The transformation of the O continuum that spaces the order statistics on O linearly is M(O), the psychophysical mapping function. This scale estimated by the order statistic is determined except for the origin and has a constant unit. It thus provides an interval scale of subjectively spaced O values if the O scale is interval or better. If a zero point can be located on the S scale that corresponds to a known point on the O scale [i.e., $M(O_0) = S_0$], then the order statistic gives a ratio scale of the S continuum.

Experiments I–III

These experiments illustrate the basic random-number-generation task as applied with some variations. The experiments are reported in chronological order and show some evolution in technique, but the findings are essentially the same in all 3.

Method

The *Ss* were asked to produce random whole numbers rapidly and with no upper limit. They were instructed not to produce long sequences of numbers in the same neighborhood and not to count or give more numbers in order than would be expected by chance. In Experiment I, instructions to this effect were given in an informal manner to 4 psychology graduate students at Johns Hopkins University. These *Ss* were friends of *E* but were kept ignorant of the intent of the experiment. Their first 26 responses were written down.

In Experiment II *Ss* were 30 male and female undergraduates at the University of California, Berkeley, who performed the task individually while waiting to serve in another, unrelated experiment. The *Ss* spoke their responses into a tape

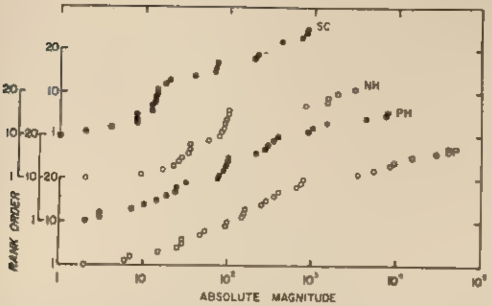


FIGURE 2. Ordinal magnitude vs. absolute magnitude for random numbers produced by 4 *S*s—SC, NH, PH, and BP: Experiment I. (Semilog coordinates.)

recorder after the following instructions were read to them:

> This is an experiment on how people think about numbers. What I would like you to do is give me any integers from one to the highest number you can imagine. I would like you to think of the numbers as going from one up, with no upper limit. Just skip about in the numbers and speak whatever number comes to mind into the microphone. Do it as quickly as you can and do not try to think about the numbers you are giving. Don't count or try to follow any kind of system. Try to make the numbers as random as you can. I will tell you when you have given me enough numbers.

Questions were answered (usually by referring to the instructions), the tape recorder was turned on, and *S* was asked to begin. Hesitancy on the part of some *S*s was overcome by repeating or paraphrasing parts of the instructions. Once *S* began giving numbers, *E* looked away and assumed a bland expression to avoid giving subtle hints about what was expected (cf., Greenspoon, 1955; Orne, 1962). The *S*s were, however, told to "be more random" or to observe no limits if they appeared to be counting or limiting themselves to numbers in a small interval, and they were urged to go faster if their rate of number production was less than 1 per 2 sec. or so. The *S*s were allowed to generate 30 numbers or more before being stopped. Numbers above 1 billion were discarded from analysis, and the first 25 acceptable numbers given were transcribed from the tape.

Experiment III was a replication of Experiment II with slightly different instructions. The *S*s were 20 male and female Pomona College undergraduates interviewed individually during breaks in an introductory psychology laboratory section. The instructions differed from those of Experiment II only in that *S*s were asked to pretend they were a "random-number-generating machine" and were asked to deliver the numbers "as if free associating." The *S*s gave about 30 responses, all of which were tape recorded. The first 15 responses under 10,000 were transcribed for analysis.

Results

The responses given by each *S* i[n] experiments were put in order befor[e] analysis was performed. Thus, i[f] sequence as generated was 3, 97, 29, it was put in order as 3, 8, 29, 97. (This is, of course, the first step in calculating the order statistics.) Numbers in the ordered sequence tend to increase exponentially in absolute magnitude as a function of ordinal position. For example, the geometric mean of the smallest number given by each *S* in Experiment II was 1.65, and the geometric mean of the largest (twenty-fifth in the order) was 3770, but the geometric mean of each *S*'s median number was 67.2. This dramatic increase in the size of the numbers with ordinal position can be seen in Figure 2, where the ordinal position of the numbers given by *S*s in Experiment I is plotted against the absolute size of the numbers.[4] The logarithmic abscissa in this figure reduces the increase of the numbers to an approximately linear one for each *S*. Two aspects of the functions of Figure 2 are of interest. First, the plots for each *S* show an overall logarithmic trend for most of their extent, but have regions of scalloping at various points. Second, the functions show some systematic departure from the logarithmic trend for numbers in the regi[on] below 10 or 15. Almost every *S* wh[o per]formed in the random-number-production task displayed these features in the numbers he gave.

[4] Intuition seems to demand that the x and y coordinates of Figure 2 be reversed so that size of an emitted number would be plotted as a function of rank order. Such a plotting would make the variable that can vary freely—size of number—the dependent variable, but this fact is not sufficient to make number the logically dependent variable. The relationship between size and rank (or between the magnitude and rank order of order statistics) resists logical division into independent and dependent variables. The arrangement of Figure 2 will be followed in subsequent figures because the function it shows is M(O). Interchanging coordinates would depict $M^{-1}(O)$ and might lead to some confusion.

The Averaging Problem

Each S's set of n random productions represents but 1 sample of the order statistics and consequently gives a noisy estimate of M(O). Averaging across Ss is clearly necessary for an accurate picture of the average M(O), but how to perform this averaging is a problem not easily resolved. The value of each of the productions obtained from any random-production experiment is necessarily on the O scale (see Figure 1). The "values" on the S scale are inaccessible, and the S continuum and its associated psychophysical function are, in fact, only intervening variables that serve to explain why subjectively equally spaced points on O are not equidistant in the units of O. In order to arrive at the best estimate of the order statistic, however, averaging of Ss' responses should take place on the S scale, because rectilinearity of sampling can hold only on the S continuum. Sampling is not rectilinear on the O scale, and the expected values of the order statistics cannot be estimated accurately by taking arithmetic means of untransformed O values. An obvious solution to the problem would be to transform the O values to numbers proportional to the S values, average these numbers, and then reconvert the averages to O values by the inverse transformation. Of course, for a precise estimate of the S_i that corresponds to each O_i, the transformation used must be M(O), and this is the unknown function to be estimated by the values of the order statistics.

Estimating order statistics with medians. Fortunately, there are a number of solutions to the averaging problem. The simplest is to take the median of the productions at each ordinal position. These are based on values on the O scale, but converting O to S by any monotonic M(O), taking medians of the S values, and then transforming the S medians back to O values by $M^{-1}(S)$ will result in O values very nearly the same as the simple median of the original O values. Thus, medians taken on the O scale can provide an estimate of the O values that correspond to the order statistics on S without requiring that any particular M(O) be assumed.

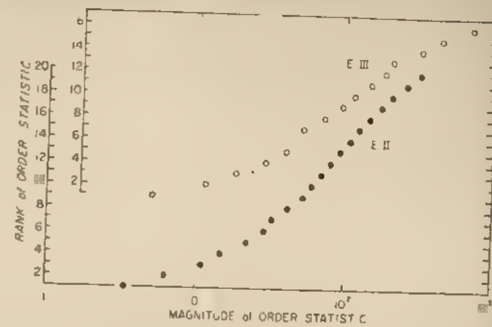


FIGURE 3. Rank of order statistic plotted against its magnitude estimated by the median of the constant-corrected numbers: Experiments II and III. (Semilog coordinates.)

The median has, however, a striking defect as a measure of central tendency in this situation. Unfortunately, with the present data, medians tend not to give an average scale but give rather the scale of the small number of "average" Ss whose productions tend to fall consistently in the middle range at each ordinal position. The medians can, however, be made much more representative of the average scale if individual differences in range and size of numbers are reduced by a correction procedure before medians are taken. An appropriate correction can allow all Ss to contribute equally to the median. Accordingly, each S's numbers in Experiments II and III were normalized by a multiplicative constant chosen so that the geometric mean of S's transformed numbers was equal to the overall geometric mean of numbers given in the experiment in which he performed. (This procedure is similar to the "modulus equalization" technique.) The constant correction was successful on at least 2 grounds: the standard deviation associated with each order statistic was reduced by approximately 20%, and nearly every S had a corrected number directly involved in the computation of one of the medians. A power transformation was also tried out as a method for normalizing the numbers, but the order statistics had standard deviations larger than those obtained with the constant transformation.

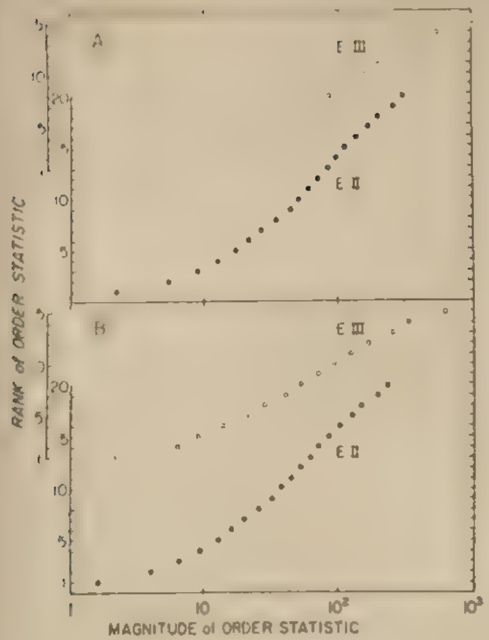


FIGURE 4. Rank vs. magnitude of order statistic for Experiments II and III. (In A the magnitude is estimated by the geometric mean of the constant-corrected numbers; in B, by the geometric mean of the untransformed numbers. Semilog coordinates.)

Figure 3 shows the relationship between the order statistics, estimated by medians of the constant-corrected numbers, and ordinal rank for Experiments II and III. The first 20 of the 25 order statistics of Experiment II and all 15 of Experiment III are shown. The general trends seen in Figure 2 show up strikingly in Figure 3. The functions are linear at the smaller numbers and show some acceleration in the semilog coordinates until a magnitude of about 12 is reached, where they become logarithmic until numbers near 1,000 are reached. Within this central range, the magnitude of the numbers from one ordinal position to the next increases by a very nearly constant percentage. For the highest numbers, the function becomes more compressed than a logarithmic function, something not seen in all the individual functions.

Estimating order statistics with geometric means. Still another averaging method is to estimate the order statistics with means of the corrected numbers at each ordinal position in the ordered output. The geometric mean seems the best one to use in this case, largely because the numbers found at each ordinal position are positively skewed, even in the corrected data. A further reason for using the geometric mean is that the functions of Figures 2 and 3 are approximately logarithmic, indicating that M(O) may be approximately logarithmic. If M(O) is a log function, taking geometric means of the O values is equivalent to taking arithmetic means on the S scale and transforming the means by $M^{-1}(O)$ to obtain estimates of values on O that correspond to the order statistics on S. If the M(O) for each *S* is a power function, geometric means will give an overall power function whose exponent is the arithmetic average of the exponents for the individual functions (Bruvold & Gaffey, 1965).

Figure 4A shows the relationship between the rank of the order statistic and the magnitude of the order statistic, estimated by the geometric means of the constant-corrected numbers for Experiments II and III. These functions show all the features of those in Figure 3, and they are somewhat smoother. In addition, they are very similar to the functions obtained by estimating the order statistics with the geometric means of the uncorrected data, as Figure 4B shows.

Cumulative Distributions of Numbers

The cumulative distribution of the numbers produced by *S*s is a concise summary of the data, although such a distribution does not bear a simple relationship to the order-statistic analysis. Table 1 shows the cumulative proportional distribution of numbers given in Experiments II and III. Table 2 shows the results when steps are taken to reduce individual differences in the numbers given. For this table, only the first 10 numbers given below 1,000 were taken from each *S*'s protocol. The numbers were then normalized by a multiplicative constant that equalized the geometric mean of *S*s' 10 numbers.

TABLE 1

CUMULATIVE PROPORTIONAL DISTRIBUTION OF NUMBERS: EXPERIMENTS II AND III

Numbers [illegible]	Cumulative proportion	
	Experiment II	Experiment III
10	[illegible]	[illegible]
20	[illegible]	[illegible]
30	[illegible]	[illegible]
40	[illegible]	[illegible]
50	[illegible]	[illegible]
60	[illegible]	[illegible]
70	[illegible]	[illegible]
80	[illegible]	[illegible]
90	[illegible]	[illegible]
100	.676	.719
110	.704	[illegible]
120	.730	[illegible]
130	[illegible]	[illegible]
140	[illegible]	.769
150	[illegible]	[illegible]
160	.751	[illegible]
170	[illegible]	[illegible]
180	[illegible]	[illegible]
190	[illegible]	[illegible]
200	[illegible]	[illegible]
300	[illegible]	[illegible]
400	.803	[illegible]
500	[illegible]	[illegible]
600	[illegible]	[illegible]
700	[illegible]	[illegible]
800	[illegible]	[illegible]
900	[illegible]	[illegible]
1,000	[illegible]	.940

These distributions show that small numbers are more likely to be selected than large ones. Such a bias in the selection of numbers is to be expected from the compressive scales of number shown in Figures 3 and 4. As the subjective distance between arithmetically equidistant numbers decreases, the numbers *Ss* select will be spaced farther apart, and thus fewer numbers will be selected from an interval of a given width.

The cumulative distributions emphasize a hypothesis about the results that will be discussed later; viz., that *Ss* perceive numbers in a linear manner, but sample from them nonlinearly. This hypothesis runs counter to the hypothesis maintained here—that *Ss* sample rectilinearly from a nonlinear subjective representation of the numbers—but it is impossible to decide in favor of one or the other on the basis of random productions alone. Experiment IV was designed to decide between these hypotheses, and the results of other experiments below are relevant to the decision between them.

Discussion

Under the assumption that *Ss* generate random integers by selecting quantities uniformly from a psychological number space (the S continuum), the results of Experiments I-III indicate that this space is compressive: successive numbers must lie closer and closer together as they become larger. Figure 5 shows values on the O dimension corresponding to the order statistics on S that would be obtained by uniform sampling from the S scale with various M(O)s.

The function labeled L would result if the psychological distances between numbers were proportional to their logarithms. According to this distribution, the subjective ruler is calibrated in the same way as the C or D

TABLE 2

CUMULATIVE PROPORTIONAL DISTRIBUTION OF NUMBERS LIMITED AT 1,000 AND ADJUSTED BY MULTIPLICATIVE CONSTANT TO EQUATE GEOMETRIC MEAN NUMBERS GIVEN: EXPERIMENTS II AND III

[illegible]	Cumulative proportion	
	Experiment II	Experiment III
1	[illegible]	[illegible]
10	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
40	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
60	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
80	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
100	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
1,000	1,000	[illegible]

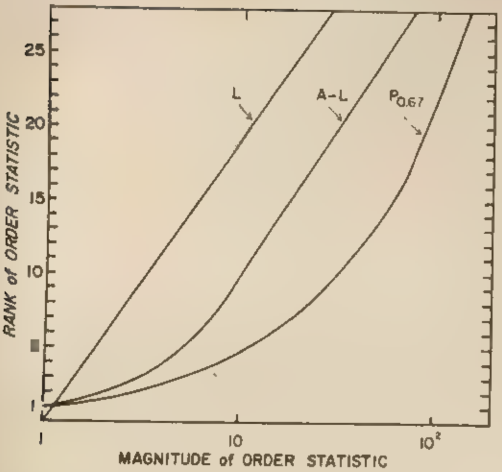


FIGURE 5. The rank vs. magnitude functions that would result for the order statistics if rectilinear samples were taken on a logarithmically compressed continuum (L), a partly logarithmic, partly linear continuum (A–L), and a continuum compressed according to a power law with an exponent of .67 ($P_{.67}$).

scale of a slide rule; that is, $M(O) = k \log O$. Numbers selected randomly from such a ruler and put in order will show an exponential increase with ordinal position, and the rank of the order statistics and the logarithm of their absolute size will be linearly related. The slope of the function is determined by the length of the ruler over which the sample is taken. The slope is not, however, at issue here. The point is that a logarithmic M(O) leads to a linear plot of the ordered samples in semilog coordinates, and numbers obtained from *Ss* in the number-generation task give such a plot for much of their range. That the obtained plots are not strictly linear in semilog coordinates is clear, and other assumptions about the psychological scale of number come closer to representing the full range of data.

The function labeled A–L in Figure 5 shows what happens if it is assumed that the psychological scale is linear for the numbers 1–10 and logarithmic thereafter. This assumption is not intuitively unreasonable, although it is mathematically cumbersome. We have a good understanding of the integers under 10 and of the relative distances among them, but above 10 or so the numbers become less distinct, and we are more willing to accept a percentage error of "rounding off" when quantities are expressed in terms of these numbers. If, in fact, we demand amount accuracy for small numbers but prefer percentage accuracy for larger numbers, a subjective scale of number like the one leading to the A–L function is implied.

A scale of numerical magnitude that has excellent empirical support is the power function (e.g., Curtis, 1970; Curtis & Fox, 1969). The $P_{.67}$ function in Figure 5 is based on a scale of number constructed according to a power function with an exponent of .67. For $P_{.67}$, $M(O) = k \cdot O^{.67}$. Numbers on this scale are spaced in such a way that their $\frac{2}{3}$ powers, if written next to them, would be linearly spaced. This power function results in order statistics that closely parallel the functions of Figures 3 and 4 for numbers below 500 or 1,000, but at the upper end of the range, the divergence of the empirical functions from a power function is great.

Fitting any function to the order statistics obtained from the number-production task is complicated by the flattening of the plots for numbers above 500 or 1,000. This deceleration would seem to indicate that the subjective scale of number becomes very compressed around 1,000 and is even more compressed for higher numbers. This apparent compression may simply result from a very poor understanding of large numbers on the part of *Ss*. The *Ss* tended, for example, to give numbers such as a billion, a trillion, and a zillion (or a quadrillion, etc.) with approximately equal frequency. It is likely that numbers of this size, and perhaps those as small as 1,000, are much less meaningful to *Ss* than smaller numbers, and they emitted the names of these numbers only in order to touch upon the vaguely understood realm of very large numbers in their output.

It is clear that numbers that are really only names and do not stand for magnitudes should be eliminated from consideration in attempts to discover the best M(O). Not only do the meaningless numbers put an unrealistic compression at the top of the scale but they also affect the function for lower numbers by entering into the averages that estimate many of the order statistics. In an attempt to eliminate the effect of large numbers on the estimate of M(O), the data of Experiments II and III were retabulated with numbers above 1,000 excluded. This procedure arbitrarily assumes that numbers become meaningless at about 1,000. The choice of 1,000 or any number is arbitrary because the lower limit of "meaninglessness" in numbers is difficult to define, and *Ss* probably differ widely in the range of numbers for which they have an intuitive understanding. Whether or not 1,000 marks the

limit of meaningful numbers, however, it is at the upper limit of numbers normally given by *S*s in magnitude-estimation experiments and thus, at the upper end of the subjective ruler that the order statistic is intended to calibrate.

Figure 6 shows the relationship between the order statistics estimated by the geometric mean of constant-corrected numbers and the ordinal rank of the statistics in Experiments II and III, when all numbers above 1,000 are eliminated from *S*s' output before the numbers are put in order for averaging. Only 10 order statistics were computed for this figure because the upper limit of 1,000 reduces the number of acceptable numbers that can be obtained from every *S*. Also plotted in Figure 6 is the function given by 5 male Johns Hopkins undergraduates who were asked, while waiting to serve in another experiment, to give random numbers, with the restriction of an upper limit of 1,000. They were allowed to emit 26 numbers, and all 26 were used. The order statistics were estimated by the geometric means of the untransformed numbers.

This figure is plotted in log–log coordinates to show that a power function describes the results. In log–log coordinates, a power function is a straight line whose slope is equal to the exponent of the power function. The functions of Experiments II and III in Figure 6 are fairly linear, although they have some curvature, especially at the upper end. Evidently, for at least some *S*s, 1,000 is well into the range of meaningless numbers. The linear portion of the curves is best described by a power function with an exponent of .60 for Experiment II and of .67 for Experiment III. The M(O) may, therefore, be a power function with an exponent around $\frac{2}{3}$. Interestingly, this value is close to being an average of the various exponents for number that Curtis and his colleagues (Curtis, 1970; Curtis & Fox, 1969) have found in fits of their 2-stage model to judgments of differences between psychophysical quantities. It is also very close to the exponent of .63 for number that Rule and Curtis (1973) found with nonmetric scaling. The fact of individual differences in output exponents (Curtis, Attneave, & Harrington, 1968) and task variables that may alter the output exponent (Poulton, 1968) indicates that the value of $\frac{2}{3}$ can be at best a first-order approximation to the exponent.

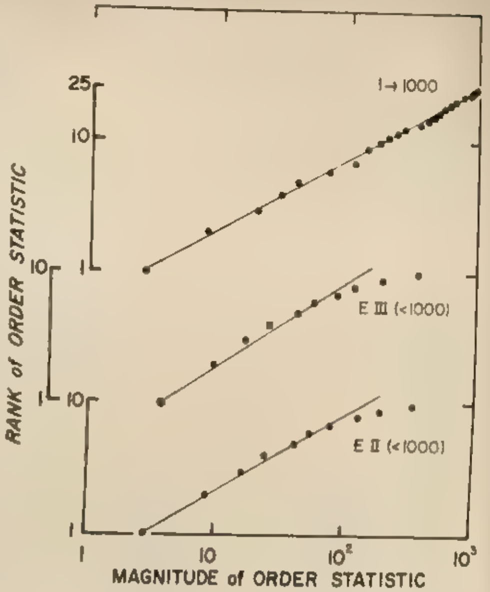


Figure 6. Rank vs. magnitude of order statistic for *S*s who generated integers with an upper limit of 1,000, labeled 1 → 1,000 and for *S*s of Experiments II and III when numbers above 1,000 were excluded, labeled E II (<1,000) and E III (<1,000), respectively. (Log–log coordinates.)

Experiments IV–VII

The results of the random-number-generating task have been taken as evidence for a psychological scale of number that is negatively accelerated. Two functions relating perceived size of number to absolute magnitude have been considered—a power function and a partly linear, partly logarithmic function (A–L). Both functions cover the scale up to about 1,000, where numbers may become meaningless, but this limit is above the upper limit of numbers commonly used in magnitude estimation.

While deciding between the power and A–L functions is a theoretically important task, determining that the obtained functions reflect the subjective scale of number and not some irrelevant factor in the number-generation task is a more immediate problem. There are a host of alternative explanations for the results, and the following experiments are directed at some of them.

Experiment IV tests the notion that the numerical continuum is subjectively linear and that nonuniform sampling results in the obtained functions. The method used allows linear sampling from a compressed "number line" to be distinguished from nonlinear sampling along a linear continuum.

Experiment V investigated the plausibility of a simple number-generation strategy that would lead to results of the sort obtained. Using this strategy, *S* would generate random single digits and decide, more or less randomly, whether to cut the string of digits at a 1-, 2-, 3-, or 4-digit or longer number. If each length of string were used equally often, about as many numbers would be produced among the units, tens, hundreds, and so on, and the ordered output functions would be approximately linear in semilog. In this experiment, *S*s were asked to give random numbers from the range ≥ 100. If *S*s were simply cutting digits into strings of random length for output, the interval 100–999, which contains numbers that are all of the same length, should be equally represented from end to end and should show a linear increase in magnitude as a function of ordinal position in the ordered output. This experiment also tests a "least effort" hypothesis of number generation. According to this hypothesis, *S*s find it easier and quicker to give small numbers. They therefore concentrate on the smaller ones, but throw in large numbers from time to time to satisfy the somewhat vague requirements of the task. Such a least effort strategy should produce a linear function over the range 100–999, since all these numbers are equally troublesome to produce.

Experiment VI investigated some effects of reinforcement and instructions on the distribution of numbers obtained in the random-number-generation task. Subtle cues from *E* can have a strong effect on behavior in a situation as unstructured as the number-generation task (Greenspoon, 1955; Orne, 1962), and it is possible that *E* was unconsciously cuing *S* to produce a distribution of the desired form. In Part 1 of this experiment, the effect of deliberately asking for larger or smaller numbers was determined. In Part 2, *S*s were presented with either an arithmetic or a geometric series (in random order) as a "good example" of a random series, and then produced their own set of numbers. If subtle cues have any effect on *S*s' number-generation behavior, these explicit instructions should have a strong effect.

Experiment VII tested the notion that, in generating their numbers, *S*s merely reflect the "spew" principle (Underwood & Schulz, 1960) by producing numbers in proportion to the frequency with which they encounter them in daily life. Small numbers are used more frequently than large ones in almost every application of numbers—from street addresses in *American Men of Science* to physical measurements (Benford, 1938; Raimi 1969)—and the decrease in usage of numbers is a logarithmic function of their size. This decrease, sometimes called Benford's law, is a sort of Zipf's (1949) law for numbers. In this experiment, *S*s were asked for random numbers with zero as the lower limit, and they produced fractional and decimal numbers in addition to integers. Fractions and decimals are probably less common in everyday occurrence and should appear less often than integers in the output if *S*s' numbers merely reflect their daily experience with numbers. Furthermore, if the scale for number is not changed when nonintegers are permitted, an explanation solely in terms of the spew principle is impossible because of the relatively low language frequency of any arbitrary noninteger.

Experiment IV

Method

The 20 *S*s who participated in Experiment III were used. After generating the random numbers in that experiment they were shown a graph like that of Figure 7A drawn on a $8\frac{1}{2} \times 11$ in. sheet. They were told that this picture expressed their instructions for the number-generation task: the abscissa represented the numerical continuum, and the horizontal line indicated that their sampling procedure gave every number an equal chance of being selected. (Every *S* but 1, a physics major who pointed out that only integers were sampled,

accepted this as an adequate description of what he did.) They were then told that they should use a new sampling rule and were shown a graph similar to Figure 7B and 7C. According to this new rule, *S*s were to sample numbers randomly, as before, but were to limit the range of numbers from which they sampled.

Every *S* generated numbers in 6 intervals defined in terms of figures like 7B and 7C drawn on $8\frac{1}{2} \times 11$ in. sheets. The intervals were of 3 widths, "narrow" (1.91 cm.), "medium" (3.82 cm.), and "wide" (7.64 cm.), and were placed with their midpoints at 1 of 2 positions on the 20-cm. abscissa (number line). One position for the intervals was described as being "among the smaller numbers" and was 7 cm. from the origin. The other was 14 cm. from the origin and was described as being "among the larger numbers." Thus, the instructions for generating numbers in, for example, the 3.82-cm. interval with its midpoint 14 cm. from the origin were, "Now I would like you to give me random numbers from an interval of *medium* width among the *larger* numbers." (The *S* was shown the drawing to illustrate the instructions and had it before him while giving numbers.) The instructions for all 6 width–position combinations were constructed according to this formula.

The *S*s often asked such questions as "How wide is 'wide'?" and were always answered with "That's completely up to you." Half of the *S*s first generated numbers from the 3 intervals placed at the smaller end of the line, then from the 3 intervals placed at the higher position. The other half generated larger numbers first, then smaller ones. At each position on the number line, all *S*s generated numbers first from the narrow interval, then from the medium-sized interval, and then from the widest one. Each *S* gave 10 numbers for each interval presented before going on to the next one.

Of course, neither the numerical span the intervals were to contain nor the magnitude of the midpoint of the intervals was specified in the instructions. This was the whole point of the experiment: *S*s themselves specified both the width and the midpoint of the intervals, and the question of interest was whether they felt an interval of a given width encompassed more numbers if it was placed at the high end of the number range than if it was placed at the low end. Since all *S*s were told that the abscissa was open ended (i.e., that it went to infinity off the page), any compression they assumed in the mapping of numbers on this line resulted from their own visualizations of the continuum and was not forced on them by the instructions.

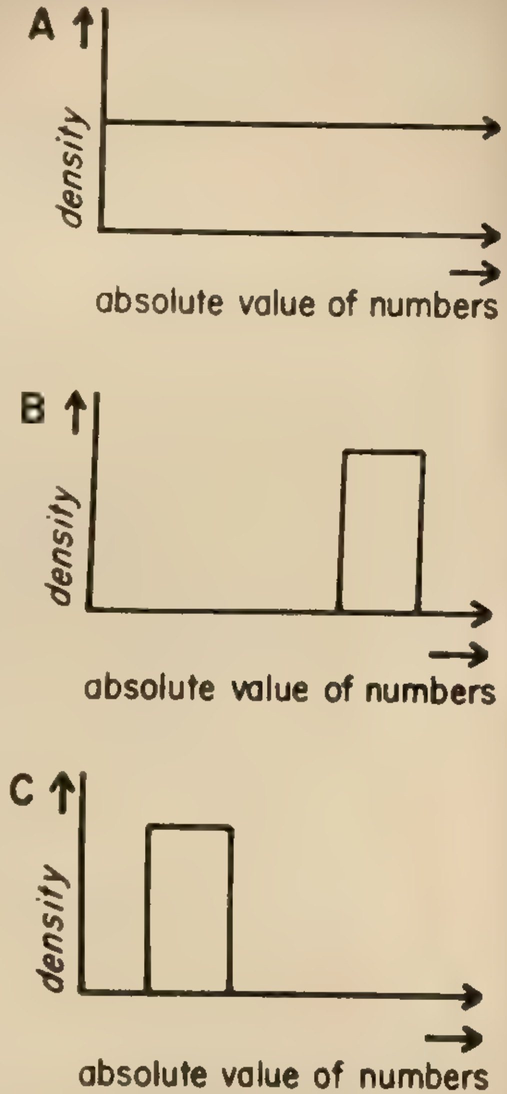


FIGURE 7. Examples of drawings used to instruct *S*s in Experiment IV.

Results

As Figure 8 shows, *S*s assumed that a given interval width, whether narrow, medium, or wide, spanned a greater numerical range when it was centered among the higher numbers than when it was centered among the smaller numbers. For example, when *S*s saw the wide interval placed among the smaller numbers, they assumed it spanned a range of 70, but among the larger numbers, the span jumped to 155. All but 3 of the 20 *S*s followed the trends shown in Figure 8, and no consistent differences between *S*s in the 2 counterbalancing conditions could be found.

An *S*s × Interval Width × Large vs. Small Numbers analysis of variance was performed on the logarithms of the range

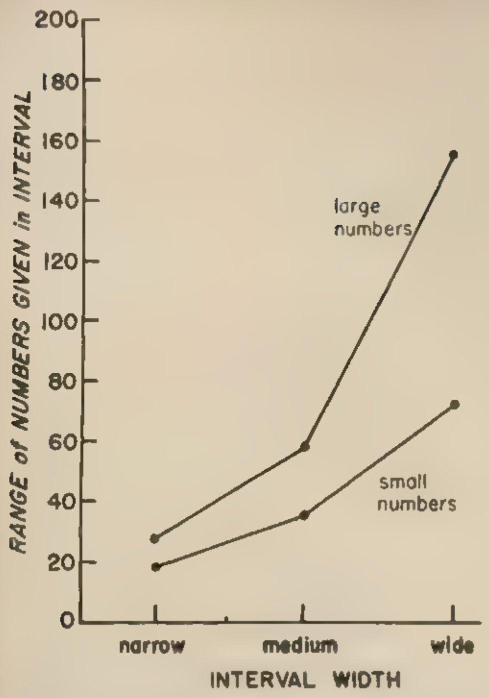


FIGURE 8. Results of Experiment IV. (Whether an interval was called narrow, medium, or wide, Ss assumed it spanned a greater numerical range when it was described as being "among the larger numbers.")

scores obtained from each *S*. The log transformation was used to reduce the differences in variability among the cells of the experiment. The large vs. small numbers variable was significant, $F\ (1, 19) = 12.7$, $p < .01$, when tested against its interaction with *S*s.

EXPERIMENT V

Method

Twelve *S*s from the population sampled in Experiment II were given the same number-generation task administered in that experiment, except that they were allowed to give only 20 numbers before being stopped. After giving the 20 random numbers, they were asked to do "the same thing again" but to maintain a lower bound of 100 instead of 1. The *S*s were urged to emit their numbers as quickly and as randomly as before, but were reminded that they should sample from "100 on up" this time. Approximately 25 numbers given by each *S* were tape recorded.

Results

Numbers over 20,000 were rejected from analysis, and the first 19 numbers given under this limit were transcribed from the tape. Three *S*s had to be rejected: 2 because they apparently interpreted the instructions as giving a lower limit of 1,000, and 1 because he gave only 12 numbers under the limit of 20,000 and would have reduced the range over which the data could be averaged. The responses of the remaining 9 *S*s were put in order, and geometric means were taken at each ordinal position. Figure 9 shows how the magnitude of the order statistics (geometric means of untransformed numbers) is related to their rank order. The results of both tasks performed by *S*s in this experiment are plotted so that the results found with the 2 ranges can be compared. The *S*s gave results similar to those of Experiments II and III when the lowest acceptable number was 1.0. Imposing a lower limit of 100 produced an almost perfectly logarithmic function, with no curvature at the lower limit. It would seem from this result that the curvature at the bottom of the other random-production scales of number is not an "end effect," but rather reflects the form of the subjective scale for small numbers.

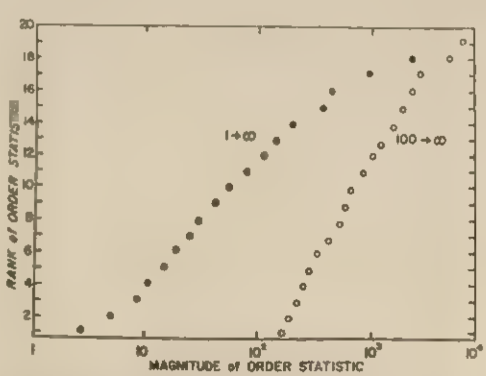


FIGURE 9. Rank vs. magnitude of order statistic when *S*s could give an unlimited range of integers ($1 \rightarrow \infty$) and when they could not give integers below 100 ($100 \rightarrow \infty$): Experiment V. (Semilog coordinates.)

Experiment VI, Part 1

Method

Experiments VI and VII were run concurrently with a total of 50 male and female *Ss*, all Pomona College undergraduates, none of whom had served in similar experiments previously. The *Ss* were assigned in rotation to the experiments in the order in which they arrived. The first 2 *Ss* were assigned to Experiment VI, Part 1, the next 2 *Ss* to Part 2, the next *S* to Experiment VII, and so on.

A total of 20 *Ss* served in Part 1 of Experiment VI. All *Ss* were initially given instructions for generating random numbers similar to those used in Experiment III, and all responses were tape recorded. After 10 responses had been given, half of the *Ss* were told that they were giving too many small numbers and should try to give more large ones. The other half were asked to give more small numbers because their numbers were too large. Some *Ss* were allowed to emit more than 10 numbers before the biasing instructions were given because they were in the process of giving a series of large (or small) numbers at the tenth item. The *Ss* in the *larger* biasing condition had to give at least 1 response under 10,000 after the tenth response before the biasing instructions were given, and *Ss* in the *smaller* condition had to give at least 1 response over 30 after the tenth before instructions were given.

Results

Figure 10A shows the order statistics (geometric means of the constant-corrected numbers) at each ordinal position before biasing instructions by *Ss* in the *larger* and *smaller* biasing conditions, respectively. In Figure 10B, the results of the bias toward larger and smaller numbers, respectively, are seen. The instructions in both cases had the intended effect of raising or lowering the overall range of numbers emitted, but the general form of the function seems little altered by the instructions. Both groups show curvature near the bottom and top of the range, with a logarithmic section in the middle, but *larger* instructions led to a more compressed scale than *smaller* instructions, especially at the top of the range. To a poor first approximation, the *larger* order statistics equal the squares of the *smaller* order statistics of the same rank.

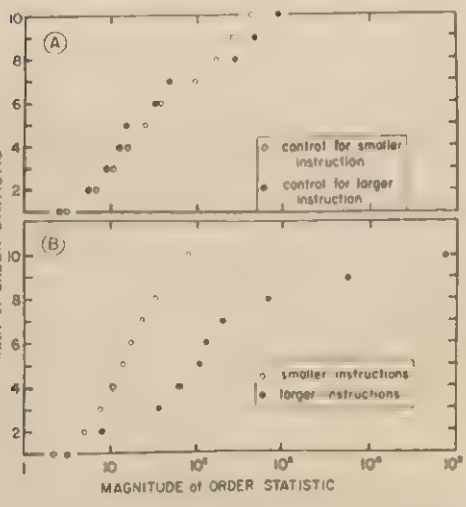


FIGURE 10. Rank vs. magnitude of order statistic before biasing instructions were given (A) and after *Ss* were told to give more small numbers (open circles) and more large numbers (filled circles) (B): Experiment VI, Part 1. (Semilog coordinates.)

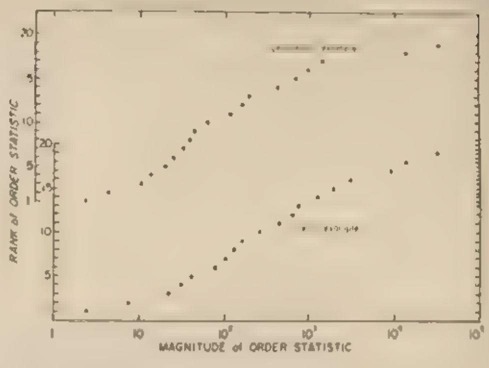


FIGURE 11. Rank vs. magnitude of order statistic for *Ss* who had heard a geometric "good example" of random numbers and for *Ss* who heard a linear "good example": Experiment VI, Part 2. (Semilog coordinates.)

Experiment VI, Part 2

Method

Twenty Pomona College undergraduates served as *Ss* in this experiment. All were given random-number-generation instructions. The *E* then played a tape-recorded series of integers, with the explanation that the series was a "good example" of what was expected of them. Two taped series were used. One contained 20 numbers in a linear series extending in equal steps from 1 to 130,000. The linear series was presented in random order, with small quantities added to or subtracted from each member to conceal the orderliness of the numbers. The other contained 20 numbers from a geometric series extending from 1 to 40,000. (The difference in the upper limits of the 2 series is the result of an error discovered too late. The effect is that the geometric

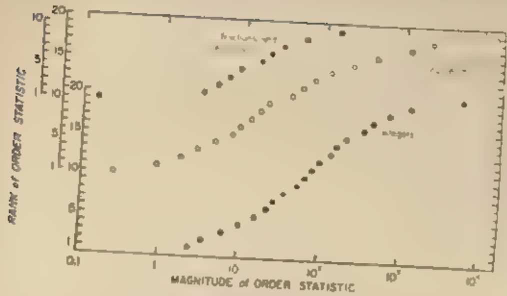


FIGURE 12. Rank vs. magnitude of order statistic when only integers could be produced and when the same *Ss* were later allowed to give fractions and decimals in addition to integers: Experiment XII. (Function for the fractions and decimals alone is also shown. Semilog coordinates.)

series is missing 3 or 4 members at its upper end.) The geometric series was also presented in random order, after small quantities were added to or subtracted from each member. Ten *Ss* had the arithmetic random series played to them, and 10 *Ss* heard the randomized geometric series. After hearing the series appropriate to his condition, each *S* dictated 20 numbers into the tape recorder.

Results

Figure 11 gives the order statistics (geometric means of constant-corrected numbers) at each ordinal position for both conditions. The random series derived from an arithmetic progression implies a subjective ruler graduated linearly, but it seemed to induce number-generation behavior consistent with the same compressed subjective ruler seen in Experiments I–III. The geometric random-number example is consistent with a subjective ruler graduated logarithmically. The numbers *Ss* produce after hearing this example imply a slightly more compressed psychological scale of number than do the numbers given by *Ss* who heard the arithmetic example, but this difference may simply be the result of the different ranges existing in the 2 examples. For most of its extent, however, the subjective scale derived from numbers given after either example of a good series is of the same form as that found in the other random-number-generation tasks.

EXPERIMENT VII

Method

In this experiment 10 *Ss*, all Pomona College undergraduates, had random-number-generation instructions read to them and sponses which were tape recorded asked to do "the same thing" aga that this time they could give frac and numbers of any size (above zero and fractional parts in addition to wh All responses were tape recorded, *Ss* to give 20 numbers, and all respons scribed from the tape and put in order

Results

Figure 12 shows the relation of the rank and the absolute size of the order statistic before and after the instructions allowing fractions were given. The function covers a wider range when fractions are allowed, but the form of the psychological scale of number implied is little alter Figure 12 is a plot of the ord based *only* on the nonintegers though 67% of the numbers nonintegers, 2 *Ss* gave only a few fractions and decimals and could not be counted in the curve for fractions and decimals only. Of the remaining 8 *Ss*, 2 *Ss* gave only 10 nonintegral numbers, and the resulting function therefore gives the results for only 10 order statistics for the 8 *Ss*. The functions in Figure 12 show that the M(O) for numbers does not depend critically on the use of integers. Furthermore, the function for nonintegers alone is very similar in form to that for integers.

Discussion: Experiments IV–VII

Experiment VII eliminates several interpretations of the results of Experiments I–III based on the spew principle. It cannot be maintained that *Ss* merely reflected their frequency of experience with numbers in the numbers they gave. Integers are probably more frequent in everyday life than with decimal parts, yet whole numbers were outnumbered 2 to 1 by fractional numbers in the responses when fractions were allowed. Except for a few common decimal numb such as 3.1416 and 1.414, and the fractions ½ and ¼, the fractional and decimal numbers *Ss* gave seemed very capricious, arbitrary, and uncommon; yet the psychological scale of number derived from only the fractions given by *Ss* is virtually the same as that found when only integers are allowed.

A modified form of the spew principle could be considered, according to which *Ss* feel

there are somehow "more" small numbers—whether small integers or small fractions. This version of the law reduces to the assumption that *S*s sample nonlinearly from a linear numerical continuum. Experiment IV shows, however, that *S*s assume that a given interval, whether narrow, medium, or wide, must span a greater numerical range as its mean increases. Such an assumption would not be made if the numerical continuum were psychologically linear.

Still another line of reasoning would have it that the numbers *S*s generate simply obey Benford's "law of first digits" (Benford, 1938; Raimi, 1969). Benford's law is a generalization based on the analysis of numbers found in a great many scientific and statistical tables. According to this law, the probability that a randomly selected number begins with a digit D is log (D + 1) − log D.

If *S*s simply use or obey this law in generating their numbers, the derived subjective scale for any decade interval will be logarithmic. If each decade interval from the units to, say, the ten thousands is sampled equally often by *S*s, then the subjective scale derived from numbers generated in the entire range of 1–99,999 will be logarithmic. While the obtained numbers are not perfectly consistent with a logarithmic subjective scale and are perhaps better described with a power law, the law of first digits comes very close to providing an explanation of the data.

A test of the first-digits artifact can be performed by examining the relative frequencies of numbers in a range wherein all begin with the same digit. If this artifact were solely responsible for the obtained results, the numbers given in any such range should be evenly distributed. Table 3 shows how numbers are distributed in the ranges 10–19 and 100–199 for Experiments II and III. All numbers given by all *S*s were included in the count, not just the first 25 or 15 used for the order statistics. Both cumulative distributions deviate significantly from linearity according to the Kolmogorov–Smirnov 1-sample test (Siegel, 1956).

In fairness to Benford (1938) it should be said that he felt that his law was really a reflection of a more general law of nature. Nature, it seems, counts exponentially or, in the terms of this study, generates quantities that are spaced linearly on a logarithmic scale, not a yardstick. Thus, the law of first digits is only 1 consequence of this law of nature, and numbers within any interval examined should be spaced with progressive compression of the scale toward the top. If this more general law is applied to behavior in the number-generation task, however, it reduces to the same assumption being made here and cannot be construed as an explanation in terms of an artifact. The *S*s agree with nature, so to speak, in using a negatively accelerated ruler to measure things. The subjective ruler probably became graduated as it is by *S*s' experience with quantities, but it cannot be argued that the effects of this experience appear in *S*s' number-generation behavior in unmediated form. The *S*s do not simply produce numbers with frequencies proportional to their experience with them, nor do they sample with a bias toward small numbers from a linear continuum. Rather, it seems that they have a nonlinear analog conception of numbers from which they sample as linearly as they can.

The other artifacts considered in Experiments IV–VII are more easily dismissed. Experiment VI showed that subtle reinforcements provided by *E* (Greenspoon, 1955) or, to phrase it differently, the "demand characteristics" of the task (Orne, 1962) can influence the range of numbers over which *S*

TABLE 3

Distribution of Numbers Given in Intervals 10–19 and 100–199: Experiments II and III

Number or interval	Frequency		
	Experiment II	Experiment III	Cumulative
Interval 10–19			
10	7	12	19
11	6	7	32
12	10	13	55
13	9	10	74
14	2	9	85
15	5	7	97
16	5	9	111
17	7	8	126
18	3	5	134
19	2	12	148
Interval 100–199			
100–109	14	11	25
110–119	14	5	44
120–129	5	3	52
130–139	5	5	62
140–149	2	3	67
150–159	4	0	71
160–169	0	4	75
170–179	1	0	76
180–189	1	3	80
199–199	1	7	88

TABLE 4

Goodness of Fit (r^2) of Power and Log Functions to Individual Number Scales

S	All 10 order statistics			First 8 order statistics		
	Log function[a]	Power function[b]		Log function[a]	Power function[b]	
	r^2	r^2	Exponent B	r^2	r^2	Exponent B
1	.970	.878	.45	.955	.937	.61
2	.986	.931	.54	.982	.972	.65
3	.976	.943	.47	.955	.978	.56
4	.974	.955	.71	.962	.976	.82
5	.966	.972	.76	.964	.974	.82
6	.902	.996	.77	.916	.994	.76
7	.984	.918	.38	.972	.962	.48
8	.867	.990	.46	.886	.990	.45
9	.978	.960	.63	.968	.974	.71
10	.945	.974	.43	.980	.964	.43
11	.960	.980	.76	.970	.980	.79
12	.970	.949	.61	.945	.988	.73
M	.956	.954	.58	.954	.974	.651
SD			.144			.143

[a] Equation: $M(O) = A + B \log O$.
[b] Equation: $M(O) = AO^B$.

samples in generating his output and the overall compression of the scale, but they do not radically alter any conclusions about the form of the scale. Experiment V controlled for the possibility that *S*s simply produce random lengths of digit strings. This strategy could produce the results obtained if *S*s generated a roughly equal number of 1-, 2-, 3-, and 4-digit numbers. Such a strategy would, however, produce a linear psychological scale for the range 100–999, and the results of Experiment V show this not to be the case. (The distributions in Table 3 reinforce this conclusion.) Experiment V also renders untenable an explanation of the results in terms of a least effort strategy.

Individual Differences in Number Production

Since the functions for numbers reported here were obtained by averaging over *S*s, it is important to determine that individual differences in number production do not somehow conspire consistently, through the evils of averaging, to produce a group function that has little relevance to the number-generation behavior of individual *S*s. Two approaches to individual differences are reported below.

Individual scales for 12 subjects. The procedure of averaging over *S*s was used because the productions given by individual *S*s seemed noisy. It is possible, however, to achieve fairly smooth functions for individual *S*s by taking repeated observations and averaging within *S*s. To obtain these functions, 12 male and female University of Oregon students were interviewed individually each day for 5 days in a row. On each day they dictated 20–30 numbers into a tape recorder under standard number-production instructions. The first 10 numbers under 1,000 given on each day by each *S* were transcribed and put in order. Order statistics for each *S*, averaged over the 5 days, were estimated with geometric means.

Table 4 presents the goodness-of-fit measure r^2, for each *S* for the logarithmic and power M(O) functions for number. The table also lists the exponents for each *S* for the power function. Because the overall scale derived from this experiment showed the extra compression that may indicate meaningless numbers at the top end, the fitting procedure was repeated for only the first 8 order statistics for each *S*, and the results of these fits are also reported in Table 4.

It seems clear from Table 4 that either a log or a power function provides an excellent approximation to the subjective scale of number. Both account for an average of over 95% of the variance in the individual scales. Deciding between the 2 functions is, however, still difficult. When all 10 points on the scale are considered, the log function gives a better fit than the power function in 7 of the 12 cases, but the power function does better in 8 of 12 cases for the first 8 points—the points that fall on the part of the range most used in magnitude estimation.

The variability in the exponents compares favorably with the variability found in exponents of power functions fit to individual *S*s' magnitude-estimation scales. The standard deviation of about .14 is somewhat less than that found by Stevens and Guirao (1964) in a study of individual loudness functions for 11 *S*s, but it is larger than the standard deviations for brightness exponents Marks and Stevens (1966) obtained in a series of 4 experiments. It is also less than the variability found by Jones and Marcus (1961) and Bruvold and Gaffey (1965). The range from smallest to largest exponent is not an ideal measure of variability, but for some studies it is the only one available for comparison. The present ratio of about 2:1 between largest and smallest is about the same as or less than ranges found

in a number of studies reviewed by Marks and Stevens, less than the ranges found by Stevens and Guirao, and less than the ranges found by Stevens and Mack (1959), as reported by Stevens (1961a). It is greater than ranges found by Marks and Stevens and by Bruvold and Gaffey.

Individual scales from Experiment II. In a further attempt to assess the generality of the group M(O) function for individual *S*s, power and logarithmic functions were fit to the data given by each *S* in Experiment II. Fitting was done only on the numbers under 190 in order to reduce as much as possible the effect of meaningless numbers on the individual scales. Since about 70% of the numbers given and 18 of the 20 order statistics reported fall in this range, this upper limit should not distort the relationship between group and individual scales very much, at least not for the more important region.

The individual functions were quite well fit by a 2-parameter power function, $Y = AX^E$. The smallest goodness-of-fit measure was .86 (only 4 of the 30 r^2 values were below.90), and the mean was .938. A 2-parameter logarithmic function, $Y = A + B \log X$, also fit the data well, but not as well as the power function. The mean r^2 for the log function was .893, and the log function was superior to the power function for only 8 of the 30 *S*s. The goodness-of-fit measure for the log function is significantly worse than for the power function at the .01 level, $t\ (29) = 3.28$.

The mean of the power function exponents was .646, a value close to that of approximately $\frac{2}{3}$ suggested here for the power function M(O), and also close to the mean exponents reported in Table 4. There was also a high degree of uniformity among *S*s in their exponents. The standard deviation of the exponents was .105, and the range was .444–.964. The standard deviation is about the size to be expected in individual fits of power functions to magnitude-estimation data. Also, the ratio of slightly more than 2 to 1 between the largest and smallest exponents is within the range one would expect if individual power functions were fit to the magnitude-estimation scales given by 30 different *S*s.

The results of the individual curve fittings indicate that power functions fit individual number scales about as well as they fit individual magnitude-estimation scales. It therefore seems reasonable to consider the group average scale for number to have the same logical status as the group scale obtained in a magnitude-estimation experiment. The analysis of individual differences in scales derived by Bruvold and Gaffey (1965) seems as appropriate for the present number scales as it is for magnitude estimation. According to this analysis, the mean of the individual power function parameters equals the parameters fit to the data averaged over *S*s.

Experiments VIII and IX: Scaling Line Length and Duration by Random Productions

The random-production technique is not limited in principle to numbers. It can be used on any dimension whose quantities *S* can produce at will. The following experiments use this technique to scale 2 dimensions, line length and duration, which admit of easy random production. Although random-production scales for continua other than number are of interest in themselves, the present reason for generating these scales is to use them in a model, discussed below, to predict magnitude-estimation scales for psychophysical continua from their random-production scales.

Method

Random line lengths and random durations were produced in Experiments VIII and IX, respectively. Ten Pomona undergraduates served individually as *S*s in these experiments, half assigned to Experiment VIII first and then Experiment IX, and half assigned in the reverse order. Instructions were essentially the same as in Experiment III except that reference was made to random line lengths and periods of time instead of to numbers.

For Experiment VIII, *S*s were given a pencil and a roll of 5-cm.-wide adding-machine paper on a .75 × 1.5 m. desk. They were told to think of line length as a continuum extending from a point to the longest line they could conceive of and were asked to draw lines that represented randomly selected lengths on this continuum. They drew their lines from the loose end of the paper toward the roll and were allowed to unroll the paper as they needed more. They were not allowed, however, to hold the pencil down and pull the paper by beneath it. Pilot *S*s who used this strategy soon fell into frantic paper pulling, and it was decided they did not have much feeling for the line lengths they produced. After *S* had finished drawing each random line, the paper was unrolled about 20 cm. and he was asked to produce another line. All *S*s drew 20 random lines.

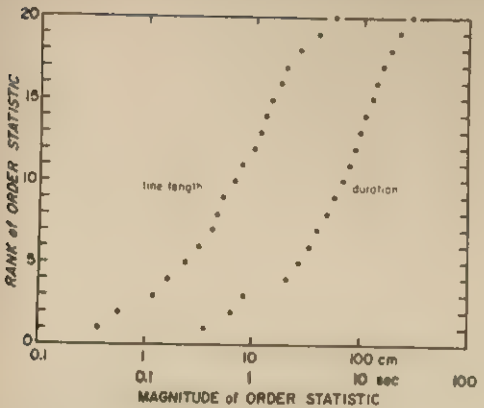


FIGURE 13. Rank vs. magnitude of order statistic for random productions of line lengths and durations: Experiments VIII and IX. (Semilog coordinates.)

In Experiment IX, *Ss* were given a microswitch that caused a Standard Electric Centisecond timer (out of *Ss'* sight) to accumulate time when it was depressed. They were told to think of durations as falling on a continuum extending from those of infinitesimal length to those as long as they could conceive of. They were asked to produce random durations from this continuum by holding the microswitch down for random periods. When *S* had completed a random duration, *E* recorded the timer reading to the nearest 1/100 sec., reset the timer, and asked for another random duration. All *Ss* produced 20 random durations.

Results and Discussion

Figure 13 shows the rank order that corresponds to each of the order statistics for both line length and duration. The functions for the 2 continua are quite similar, both being compressed and having a form that is in fact similar to that found with random productions of numbers. It is interesting that a compressed subjective scale for length is consistent with judgments and productions of sums of lines (Kreuger, 1970), even though judgments of length are linear (Stevens & Galanter, 1957). Equivalent evidence for a compressed duration scale does not seem to exist.

Randomly produced numbers have a considerably greater range than random line lengths or durations. The value of the order statistic in free number production often spans 5 log units, but the order statistics for both line length and durat[illegible] fit comfortably within 3 log units. Ma[illegible] mum quantities given by individual [illegible] also reflect this difference: the large[illegible] numbers were often in the billions or highe[illegible] but the longest random line drawn was [illegible] cm. and the longest duration produced w[illegible] just over 2 min. It is possible that [illegible] in these random production tasks feel, [illegible] they do not feel when generating numbe[illegible] that they should limit themselves to fai[illegible] small quantities. It also may be that va[illegible] magnitudes are easier said than done. [illegible] any event, the following analysis shows tha[illegible] the scales derived from the random-pro[illegible] duction tasks are not inconsistent wit[illegible] behavior seen in magnitude-estimatio[illegible] experiments.

Simulation of Magnitude Estimation and Cross-Modality Matching

A magnitude-estimation task has been characterized as one in which *S* matches numbers to the perceived size of his sensation according to a rule. If the random-production scales derived for number, line length, and duration reflect the perceived magnitudes of these dimensions, then the scales specify the rule *S* uses for matching. Magnitude-estimation behavior can thus be simulated by artificially matching numbers

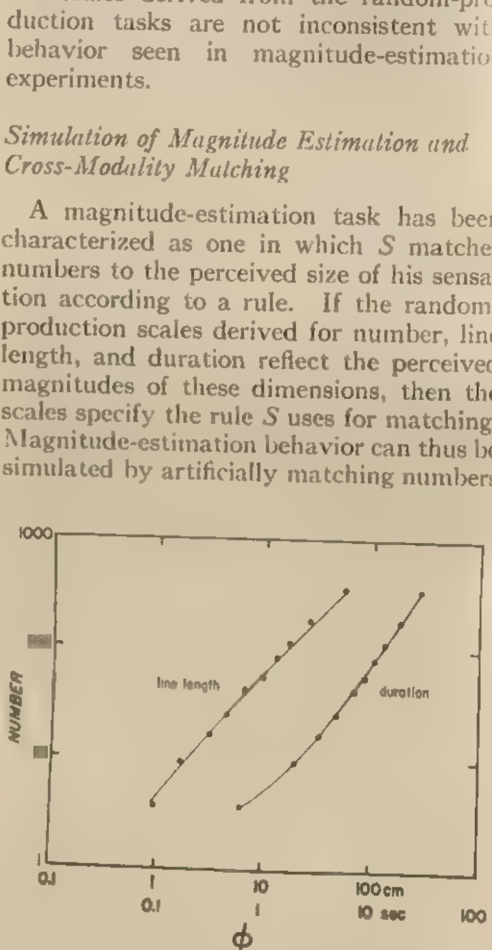


FIGURE 14. Simulated psychophysical function for the match between number and line length and between number and duration, using data from Experiments III, VIII, and IX. (Log log coordinates.)

to quantities on the estimated dimension in such a way that the subjective magnitudes of the numbers are paired with equivalent subjective quantities on the dimension. Since the order statistics are assumed to be equally spaced on the S dimension, the order statistic of Rank N on one dimension should be matched with the order statistic of the same rank on the other dimension to create the x–y pair that simulates a single magnitude estimation.

Figure 14 shows the scales that result when numbers are matched to equivalent subjective quantities of length and duration according to the random-production scales. The numbers used in these matches are from the scale of Experiment III when numbers above 1,000 were omitted (see Figure 6 and associated discussion), and the lengths and durations are from the scales of Experiments VIII and IX. Because only 10 order statistics were available for the scale of number used for the matches and 20 had been taken for duration and line length, order statistics of Rank N for number were matched to order statistics of Rank 2N for length and duration. When the random-production scale for the full set of numbers produced in Experiment III is used, functions similar to those of Figure 14 emerge, but they show acceleration at the upper end. Presumably, this acceleration results from the distorted scale for the very large numbers.

The linearity of the functions of Figure 14 in log–log coordinates indicates that power functions describe them well. The curvature in these functions occurs near the lower end and might be removed by a threshold correction. In this situation, however, the curvature probably means that the matching procedure does not line up the 2 continua in exactly the same way *S*s doing magnitude estimation would. The slope of the linear portion of the simulated apparent length function is close to 1.0, and that for the duration function is about 1.3. The simulated apparent length scale is thus quite close to that found by magnitude estimation, but the duration function is somewhat steeper than the 1.1 slope found in magnitude estimation (Stevens & Galanter, 1957). This difference between obtained and predicted slopes for duration may result from an artificial compression of the time dimension in random productions—even 1 min. of holding the button down seems like a very long time—but it may also be a result of the fact that there is no regression effect to lower slope in the simulated matches (Stevens & Greenbaum, 1966).

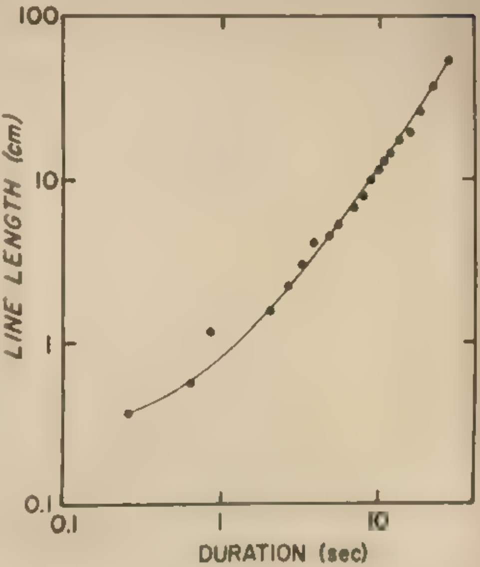


FIGURE 15. Simulated cross-modality match between line length and duration. (Log log coordinates.)

The random-production scales for line length and duration also allow the cross-modality match (Stevens, Mack, & Stevens, 1960; Stevens, 1959) for these dimensions to be simulated. Figure 15 shows the simulated result of a task where line lengths are drawn to match the apparent length of periods of time. For this simulation, order statistics of Rank N for length were paired with order statistics of Rank N for duration, and 20 simulated length–duration matches were thereby made. The function these simulated matches trace out is evidently well fit by a power function, as is shown by its near linearity in log–log coordinates. The curvature near the bottom seems to indicate a needed threshold correction, but again it may mean that match-

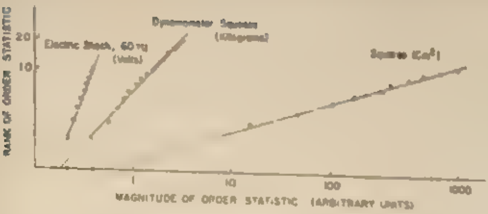


FIGURE 16. Random-production scales for subjective intensity of electric shock, force of handgrip squeeze, and areal extent, all arbitrarily placed on the X axis. (Log–log coordinates.)

ing behavior is not perfectly simulated. The linear portion of this function has a slope of about 1.4. This slope is somewhat steeper than it should be and may be too steep for a number of reasons, including the lack of a regression effect.

Random productions for other continua. Random-production scales have been obtained for subjective force of handgrip squeeze, subjective intensity of a 60-Hz. electric shock, and subjective areal extent (Banks, 1973). Figure 16 shows these scales with threshold corrections of 7 kg. for the handgrip scale and .1 cm.² for the area scale. The technique used to obtain these scales was very similar to that used for the length and duration scales presented in Figure 13. The *Ss* used a hand dynamometer to produce the handgrip squeezes, and a variable transformer, connected through a step-down transformer to saltwater finger electrodes, to produce random shocks. Areas were produced by freehand drawing of squares on unlined $8\frac{1}{2} \times 11$ in. paper.

A power function provides an excellent fit to each of these scales, accounting for over 98% of the variance of the means in every case. The exponents for the handgrip, shock, and area scales are 1.1, 2.5, and .30, respectively. These exponents are smaller than magnitude-estimation exponents for the same continua, but their small size is not consistent with a least effort (or least *pain*, in the case of shock) hypothesis of random-production behavior. First, the shock scale (which should be most compressed by such a hypothesis), is the most accelerated one, and the handgrip scale is the next most accelerated, while the area scale (whose large quanti[ties] are, if anything, easier to produce th[an] small quantities) is the most compress[ed]. Thus, *Ss* chose to give themselves more b[ig] than little shocks and more hard than eas[y] handgrip squeezes, but they drew mor[e] little squares than big ones.

Another version of the least effort hypothesis, that *Ss* produce a compressed version of the "true" psychophysical scale as given by the magnitude-estimation function, is also inconsistent with the results. The shock scale (which should be the most relatively compressed by this account) underestimates the magnitude-estimation exponent the least, the handgrip scale the next least, and the area scale the most.

It seems best to view random production as a scaling technique, especially since the least effort hypothesis of the results is disconfirmed and nothing plausible seems available to replace it. The fact that the random-production scales for shock, handgrip, and area are consistent with magnitude-estimation scales under the assumption of a nonlinear scale for number is discussed in the following section.

GENERAL DISCUSSION

Implications of scales of number. Although previous research (e.g., Curtis, 1970; Rule, 1969; Rule & Curtis, 1973) and the present evidence point to a nonlinear psychophysical function for number, the question of precisely what this function is has not been answered. The evidence seems to indicate a power function with an exponent of approximately $\frac{2}{3}$, but a logarithmic function with an inelegant linear disturbance near the bottom cannot be excluded. Both functions cover the range of numbers used in magnitude estimation well, and both show too little compression for the very large numbers.

Similar as they are in terms of fitting the random-production scales for number, these 2 functions have quite different implications for the form of the psychophysical functions of the dimensions to which numbers may be matched. If the scale for number is logarithmic and power functions are obtained in magnitude estimation, then it is unlikely that the function for the estimated dimensions is anything but logarithmic. Ekman (1964) has shown

how power functions will result for magnitude estimation if all continua, including number, have logarithmic psychophysical functions. With a logarithmic function for number, the psychophysical law for the estimated continua cannot be a power function, and whatever psychophysical law they do obey must be virtually indistinguishable from Fechner's (1860).

If, on the other hand, the subjective size of numbers grows as the $\frac{2}{3}$ power of their absolute magnitude, then the power functions found in magnitude estimation imply that the perceived magnitudes of the continua being judged also rise as a power function of the physical measure. The magnitude-estimation function will, however, have an exponent 1.5 times too large. This can be seen from the following considerations. The preceived size of numbers, by this argument, obeys the following function:

$$\Psi_n = k_n N^{\frac{2}{3}},$$

where Ψ_n is the perceived size of numbers, N is the absolute size of the numbers, k_n is an arbitrary constant, and $\frac{2}{3}$ is the exponent in the power function for apparent size of number. If the perceived size of magnitudes on any other dimension, ϕ_1, grows according to the function

$$\Psi_1 = k_1 \phi_1^{\beta},$$

where β is the exponent in the power function for the apparent size of ϕ_1, then the match between numbers and ϕ_1 will give the function

$$N = k_2 \phi_1^{\beta/\frac{2}{3}} = k_2 \phi_1^{1.5\beta}.$$

Even if all exponents from magnitude estimation misrepresent the subjective psychophysical exponent by a common factor, cross-modality matches will still be correctly predicted. Suppose, for illustration of this fact, that S's task is to match 2 continua, ϕ_1 and ϕ_2, and that magnitude estimation gives the following psychophysical power functions:

$$\Psi_1 = k_1 \phi_1^{\alpha}$$

and

$$\Psi_2 = k_2 \phi_2^{\beta}.$$

The predicted match for these 2 continua, when ϕ_1 is adjusted by S to equal the apparent size of magnitudes of ϕ_2, will be

$$\phi_1 = k_3 \phi_2^{\beta/\alpha}$$

(Stevens, 1959).

If the overestimation of the exponents due to the nonlinear scale of number used in magnitude estimation is taken into account, the exponents used for prediction should be $\frac{2}{3}(\alpha)$ and $\frac{2}{3}(\beta)$, and the predicted match should therefore be

$$\phi_1 = k_3 \phi_2^{\frac{2}{3}(\beta)/\frac{2}{3}(\alpha)} = k_3 \phi_2^{\beta/\alpha},$$

which is the same prediction obtained from the magnitude-estimation exponents.

It should be noted that few experiments employing magnitude estimation depend on the precise size of the exponent. Rather, they usually set out to prove that a power function does (or does not) fit a given estimation scale, or else they study relative values of power function exponents. Thus, the conclusions of very few of these experiments would have to be changed if the psychophysical function for number is a power function.

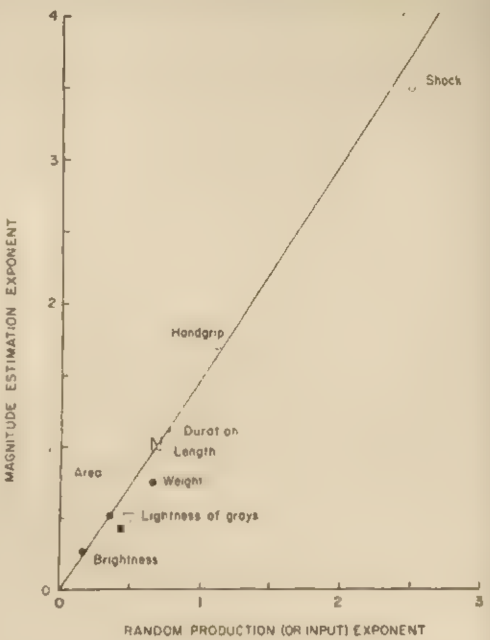


FIGURE 17. Relationship between random-production exponent (open circles) or input (sensory) exponent (filled circles) and that obtained by magnitude estimation for a number of continua. (Linear coordinates.)

Predicting Magnitude Estimation Exponents

If magnitude-estimation scales overestimate the psychophysical exponent by a factor of $\frac{3}{2}$ and if random production yields a scale for the continuum that is unaffected by the numerical response scale, then the magnitude-estimation exponent should equal 1.5 times

the random-production exponent. The degree to which this prediction holds can be seen in Figure 17. In this figure, the open circles show magnitude-estimation exponents for the continua scaled here, plotted as a function of their random-production exponents. Random-production exponents for handgrip, shock, and area were obtained from Banks (1973), and exponents for duration (.75) and length (.68) were estimated from the random-production scales of Experiments VIII and IX. The magnitude-estimation exponents are those reported in Stevens (1961a) except for area, which is an average of the values obtained by Stevens and Guirao (1964), Teghtsoonian (1965), and Curtis and Rule (1972). If the magnitude-estimation exponent is 1.5 times larger than the random-production exponent, these points should fall on a line with a slope of 1.5. As is seen, the points are well approximated by this line.

Also plotted in this figure are points showing the relationship between "input" exponents of the 2-stage model of magnitude estimation (cf. Curtis et al., 1968). These points have been taken from a number of papers on the 2-stage model (Curtis, 1970; Curtis et al., 1968; Curtis & Fox, 1969; Rule & Curtis, 1973) and they summarize results for several continua. Since the input exponent in this model is an estimate of the sensory scale of the continuum without the effect of the numerical response scale, the relationship between the input and the magnitude-estimation exponents should approximate a straight line with a slope of 1.5 and, as seen in Figure 17, it does quite well.

Finally, the N in the figure relates the subjective exponent for number, as estimated in this paper and by Rule and Curtis (1973), to the exponent found in magnitude estimations of number by Rosner (1965), Stevens (cited in Marks, 1968), and Banks (unpublished experiment, 1968). Of course, numerical magnitude estimations of numbers lead to an almost perfectly linear scale, and the N plotted in this figure serves only to determine the slope of the line on which the other points should fall.

Indirect Evidence for a Nonlinear Numerical Response Scale

Given that the psychophysical law for number is a power function with an exponent of about $\frac{2}{3}$, it still remains to be shown that this law governs the use of numbers as responses in a magnitude-estimation experiment. The main reason for studying the subjective scale of number is to determine its effect on magnitude-estimation scales. The psychophysical scales of number generated here may, however, apply only to unconstrained numbers.

The work of Curtis and his colleagues (Curtis, 1970; Curtis et al., 1968; Curtis & Fox, 1969) with the 2-stage model provides the best direct evidence that a decelerated power function of number is at work when estimates of sensory magnitude are made, but there exist some other less direct indications that this is the case. Equisection and related scaling tasks that do not give specific numerical instructions to *S* presumably operate entirely on the subjective continuum. There is no need to assume that *S*s interpret numerical instructions correctly, only that they understand the meaning of equality of sense distances (Torgerson, 1958). If the numerical response scale causes the psychophysical power function derived from magnitude estimation to have too large an exponent, then the scales from equisection and related techniques should be power functions with exponents smaller than those of magnitude estimation. This is usually the case (Stevens, 1961b, 1971a), but the fact that these techniques are subject to bias effects that can lead to internal inconsistencies and to various different predicted psychophysical functions (Garner, 1954b, 1958; Stevens, 1957) makes a more precise comparison with magnitude-estimation scales impossible.

In category scaling, *S* is instructed to space the categories evenly over the stimulus continuum. Although the instructions usually designate the categories with numbers, the task requires only that *S* break the continuum into a number of apparently equal intervals. Here, too, there are a number of sources of bias operating, and the form of the category scale is not solely determined by the individual stimulus magnitudes. Marks (1968) has shown that power functions can describe category scales obtained for a wide variety of continua if it is assumed that the average category ratings are on an interval scale and must be readjusted for unit and zero point. The parameters in the power functions Marks fits to category scales reflect the sources of bias of various experimental procedures, but the exponents always remain below those found in magnitude estimation.

Fractionation and ratio-production scales should, when power functions describe them, have exponents approximately the same as the

magnitude-estimation function for the continuum. This is so because the experimentally prescribed ratios will be interpreted nonlinearly by S, as are numbers in magnitude estimation. These techniques give the same exponent as magnitude estimation when production and fractionation are used together (Scharf & Stevens, 1961; Stevens, 1962), but, individually, the techniques give scales that seem overly subject to procedural biases (Garner, 1954b). It is of interest that Garner (1954a) was able to equate fractionation and equisection scales of loudness by assuming that Ss fractionated according to a ratio of 1:1.7 instead of the prescribed ratio of 1:2. If the power function for perceived number is $\frac{3}{4}$, 1:2 will define a subjective ratio close to 1:1.7. As a further bit of evidence, Stevens (1970, 1971b) has noted that when power functions are fit to neuroelectric correlates of stimulus intensity, the exponent is typically smaller than that for magnitude estimation of the same intensity continuum.

Finally, it should be noted that Stevens (1971a) uses the term "virtual exponent" to refer to the undersized exponents generally produced by the scaling techniques discussed above, and he provides a more thorough review of this topic. Stevens does not, however, give an explanation of why the virtual exponent should be smaller than the "actual" exponent obtained in magnitude estimation. An important advantage of the conclusion that numbers are used nonlinearly in magnitude estimation is that it predicts the general relationship between the virtual exponent and the exponent obtained in magnitude estimation.

REFERENCES

Attneave, F. Perception and related areas. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. IV. New York: McGraw-Hill, 1962.

Banks, W. P. A new psychophysical ratio scaling technique: Random production. *Bulletin of the Psychonomic Society*, 1973, **1**, 273–275.

Benford, F. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 1938, **78**, 551–572.

Bruvold, W. H., & Gaffey, W. R. Subjective intensity of mineral taste in water. *Journal of Experimental Psychology*, 1965, **69**, 369–374.

Curtis, D. W. Magnitude estimations and category judgments of brightness and brightness intervals: A two-stage interpretation. *Journal of Experimental Psychology*, 1970, **83**, 201–208.

Curtis, D. W., Attneave, F., & Harrington, T. L. A test of a two-stage model of magnitude judgment. *Perception & Psychophysics*, 1968, **3**, 25–31.

Curtis, D. W., & Fox, B. E. Direct quantitative judgments of sums and a two-stage model for psychophysical judgments. *Perception & Psychophysics*, 1969, **5**, 89–93.

Curtis, D. W., & Rule, S. J. Magnitude judgments of brightness and brightness difference as a function of background reflectance. *Journal of Experimental Psychology*, 1972, **95**, 215–222.

Dawson, W. E. Magnitude estimation of apparent sums and differences. *Perception & Psychophysics*, 1971, **9**, 368–374.

Ekman, G. Is the power law a special case of Fechner's law? *Perceptual and Motor Skills*, 1964, **19**, 730.

Ekman, G., & Hosman, B. Note on subjective scales of number. *Perceptual and Motor Skills*, 1965, **21**, 101–102.

Fairbank, B. A., Jr. Experiments on the temporal aspects of number perception. *Dissertation Abstracts International*, 1969, **30**, 403.

Fechner, G. T. *Elemente der Psychophysik.* Leipzig, East Germany: Breitkopf & Hartel, 1860.

Garner, W. R. A technique and a scale for loudness measurement. *Journal of the Acoustical Society of America*, 1954, **26**, 73–88. (a)

Garner, W. R. Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, 1954, **48**, 218–224. (b)

Garner, W. R. Advantages of the discriminability criterion for a loudness scale. *Journal of the Acoustical Society of America*, 1958, **30**, 1005–1012.

Goude, G. *On fundamental measurement in psychology.* Stockholm, Sweden: Almqvist & Wiksell, 1962.

Greenspoon, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *American Journal of Psychology*, 1955, **49**, 16–22.

Hogg, R. V. *Introduction to mathematical statistics.* New York: Macmillan, 1969.

Jones, F. N., & Marcus, M. J. The subject effect in judgments of subjective magnitude. *Journal of Experimental Psychology*, 1961, **61**, 40–44.

Kreuger, L. E. Apparent combined length of two-line and four-line sets. *Perception & Psychophysics*, 1970, **8**, 210–214.

Marks, L. E. Stimulus-range, number of categories and form of the category-scale. *American Journal of Psychology*, 1968, **81**, 467–479.

Marks, L. E., & Stevens, J. C. Individual brightness functions. *Perception & Psychophysics*, 1966, **1**, 17–24.

Moyer, R. S., & Landauer, T. K. Time required for judgments of numerical inequality. *Nature*, 1967, **215**, 1519–1520.

Orne, M. T. On the social psychology of the psychological experiment. *American Psychologist*, 1962, **17**, 776–783.

Poulton, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, **68**, 1–19.

Raimi, R. A. The peculiar distribution of first digits. *Scientific American*, 1969, **221**, 109–120.

ROSNER, B. S. The power law and subjective scales of number. *Perceptual and Motor Skills*, 1965, **21**, 42.

RULE, S. J. Equal discriminability scale of number. *Journal of Experimental Psychology*, 1969, **79**, 35–38.

RULE, S. J., & CURTIS, D. W. Conjoint scaling of subjective number and weight. *Journal of Experimental Psychology*, 1973, **97**, 305–309.

SCHARF, B., & STEVENS, J. C. The form of the loudness function near threshold. *Proceedings of the Third International Congress on Acoustics*, 1961, 80–82.

SEKULER, R., RUBIN, E., & ARMSTRONG, R. Processing numerical information: A choice time analysis. *Journal of Experimental Psychology*, 1971, **90**, 75–80.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

STEVENS, J. C., & GUIRAO, M. Individual loudness functions. *Journal of the Acoustical Society of America*, 1964, **36**, 2210–2213.

STEVENS, J. C., & MACK, J. D. Scales of apparent force. *Journal of Experimental Psychology*, 1959, **58**, 405–413.

STEVENS, J. C., MACK, J. D., & STEVENS, S. S. Growth of sensation on seven continua as measured by force of handgrip. *Journal of Experimental Psychology*, 1960, **59**, 60–67.

STEVENS, S. S. On the psychophysical law. *Psychological Review*, 1957, **64**, 153–181.

STEVENS, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology*, 1959, **57**, 201–209.

STEVENS, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication.* Cambridge, Mass.: M.I.T. Press, 1961. (a)

STEVENS, S. S. To honor Fechner and repeal his law. *Science*, 1961, **133**, 80–86. (b)

STEVENS, S. S. The surprising simplicity of sensory metrics. *American Psychologist*, 1962, **17**, 29–39.

STEVENS, S. S. Neural events and the psychophysical law. *Science*, 1970, **170**, 1043–1050.

STEVENS, S. S. Issues in psychological measurement. *Psychological Review*, 1971, **78**, 426–450. (a)

STEVENS, S. S. Letter to the editor. *Science*, 1971, **172**, 502. (b)

STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, 1957, **54**, 377–411.

STEVENS, S. S., & GREENBAUM, H. B. Regression effect in psychophysical judgement. *Perception & Psychophysics*, 1966, **1**, 439–446.

TEGHTSOONIAN, M. The judgment of size. *American Journal of Psychology*, 1965, **78**, 392–402.

TORGERSON, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

UNDERWOOD, B. J., & SHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

ZIPF, G. K. *Human behavior and the principle of least effort.* Cambridge, Mass.: Addison-Wesley, 1949.

(Received January 24, 1973)

CONTENTS—(*Continued*)

Beginning in 1975 the *Journal of Experimental Psychology* will appear as four independently edited and distributed sections:

The **Journal of Experimental Psychology: Human Learning and Memory** (Lyle E. Bourne, Jr., Editor) will publish experimental studies on fundamental acquisition, retention, and transfer processes in human behavior.

The **Journal of Experimental Psychology: Human Perception and Performance** (Michael I. Posner, Editor) will publish experimental studies designed to foster understanding of information-processing operations and their relation to experience and performance.

The **Journal of Experimental Psychology: Animal Behavior Processes** (Allan R. Wagner, Editor) will publish experimental studies concerning the basic mechanisms of perception, learning, motivation, and performance, especially as revealed in the behavior of infrahuman animals. The studies should make significant contributions to general behavior theory.

The **Journal of Experimental Psychology: General** (Gregory A. Kimble, Editor) will publish experimental articles in any of the three areas represented by the above sections when the articles involve a longer, more integrative report leading to an advance in knowledge that is judged to be of interest to the entire community of experimental psychologists. The section will include but not be limited to articles such as those that have appeared as *Journal of Experimental Psychology* monographs and as chapters in contemporary books of "advances." Republication of a limited number of data may be permitted if necessary to make the article complete and definitive.

In some instances it may be difficult for the author to decide whether one of the above listed sections or the *Journal of Comparative and Physiological Psychology* is most appropriate for a given manuscript. The author should weigh the essential implications of the research rather than the specific techniques or subject population employed. Is the central message concerned with the issues and concepts of behavior theory (*JExP*) or with the biological substrates of behavior (*Journal of Comparative and Physiological Psychology*)? Is the study primarily addressed to theory of human behavior (*JExP: Human Learning and Memory* or *JExP: Human Perception and Performance*) or to theories equally or more relevant to infrahuman subjects (*JExP: Animal Behavior Processes*)? Is the major emphasis on the acquisition, retention, or transfer of new information, concepts, or problem-solving strategy (*JExP: Human Learning and Memory*) or on the organization or control of structures or mental operations, where the acquisition of new skills is only incidental (*JExP: Human Perception and Performance*)?

Manuscripts should be sent directly to the Editor of the relevant section at the address listed below. In cases where an Editor believes a manuscript is clearly more appropriate for an alternative journal of the American Psychological Association, he may redirect the manuscript with the approval of the author. A final action taken by one of the Editors will be final for all sections of *JExP*.

Manuscripts should be submitted in triplicate and according to the instructions contained in the *Publication Manual* of the American Psychological Association.

Authors should note that limitations on total pages make it necessary to restrict publication to those papers that in the opinion of the Editors make a notable contribution to the literature. Preference can be expected for sets of interrelated experiments that are definitive or provocative.

Addresses for the 1975 Editors-Elect are as follows: *JExP: Human Learning and Memory*, Lyle E. Bourne, Jr., Department of Psychology, University of Colorado, Boulder, Colorado 80302; *JExP: Human Perception and Performance*, Michael I. Posner, before June 1, 1974, same as Allan R. Wagner, after June 1, Department of Psychology, University of Oregon, Eugene, Oregon 97403; *JExP: Animal Behavior Processes*, Allan R. Wagner, Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510; *JExP: General*, Gregory A. Kimble, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1974: $60 domestic, $61 foreign, $6 single issue. APA member rate: $25. Write to Subscription Section, APA.

Reprints: Authors may order reprints of their articles from the printer when they receive galley proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., 175 Fifth Avenue, New York, New York 10010; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. Undelivered copies resulting from address changes will not be replaced; second-class mail is not forwarded unless subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Published monthly in two volumes, six issues per volume, by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Washington, D.C., and additional mailing offices.

802(5)
2.8.74
KL

Journal of Experimental Psychology

Vol. 102, No. 3 March 1974

RHYTHMIC STRUCTURE IN AUDITORY TEMPORAL PATTERN PERCEPTION AND IMMEDIATE MEMORY[1]

PERSIS T. STURGES
California State University, Chico

JAMES G. MARTIN[2]
University of Maryland

In 3 experiments, *S*s heard continuous sequences of 14 or 16 binary elements (high- and low-buzz tones). These were either 7-element or 8-element patterns repeated once. Dummy sequences were included also to provide a *same–different* (repeating or changed) judgment task. The *S* also wrote each sequence. Half of the patterns of each length followed a rule for rhythmic structure and half did not. Of particular interest is that 7-element patterns, though shorter than 8-element patterns, cannot have a simple rhythm. In general the results were that rhythmic patterns were more easily recognized as repeating than nonrhythmic patterns, and 8-element patterns were more easily recognized as repeating than 7-element patterns. Similar results held for written recall.

Martin (1972) has presented a view of real-time information processing in which the perception and production mechanisms underlying temporally unfolding behavior are linked by the concept of rhythmic action. Very briefly, the basic notions are these. Natural constraints on human movement entail that sequences of movement-produced sounds, e.g., connected speech and music, are rhythmically patterned, i.e., they are sequences of "accented" and "unaccented" elements temporally organized *in a certain way.* A descriptive rule has been proposed for characterizing the organization of these natural patterns and shows that their temporal structure is based upon relative timing constraints which go beyond the simple ordering (concatenation) of elements. The rule also shows the way in which rhythmic patterns have a hierarchical structure.

Of relevance here is the relation between characteristics of a stimulus, in this case auditory temporal patterns, and the way the stimulus is processed (Garner, 1970b). If auditory stimulus sequences normally occurring in the natural environment are characteristically organized in a certain way, e.g., they are rhythmic, then one might reasonably expect that perceptual mechanisms as they have evolved will be biased to listen for sounds organized in this certain way. This basic notion underlies the present work.

If an auditory pattern is rhythmic, one of its properties is that certain elements in the pattern, main accents in particular, are temporally redundant. The potential

[1] This research was supported by U.S. Public Health Service Grant MH-16726 from the National Institute of Mental Health to the second author, and by a grant from the Biomedical Sciences Support Committee, University of Maryland, to the Center for Language and Cognition. Nancy Hurlbut, Dianne Reamy, Judith Sprei, and Judith Zachai helped with the experiments. Experiment III was reported at the meeting of the Psychonomic Society, St. Louis, November 1971.

[2] Requests for reprints should be sent to James G. Martin, Department of Psychology, University of Maryland, College Park, Maryland 20742.

consequence for perception of such an organization of the input sequence is that once early accents in the pattern are heard, later accents, and the end of the pattern as well, can be anticipated in real time. The question of interest then is whether perceptual mechanisms are in some way designed to make active use of these temporal redundancies in the pattern during the on-line input of the pattern itself. Alternatively, one could postulate a mechanism that treats the sequence of elements in a pattern as a concatenated string, so that even if there are temporal redundancies in the sequence these redundancies are ignored during listening and the elements registered one by one as they are heard, in left-to-right fashion. A further alternative, if elements occur at rates too fast to be handled sequentially, as in speech (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), is to suppose that processing of the sequence consists in first accepting a whole string of elements, e.g., those making up a phrase or other "decision unit" (Miller, 1962), and then processing the whole string simultaneously upon reaching the end of the unit, the occurrence of which has just been signaled by a pause or other terminal cue. This alternative also does not take account of any temporal redundancies in the sequence, and, like the analysis-by-synthesis (Halle & Stevens, 1964) and similar models, would appear to require extremely rapid processing between decision units. By contrast, the present theory suggests that the mechanisms for on-line listening in real time are optimally designed to make use of the temporal redundancies always present in humanly produced, i.e., rhythmically structured, sound inputs. The redundancies are highly correlated with the distribution of information (melodic or linguistic) in the sequences and hence are not ignored but are made use of during ongoing perception in a highly efficient way (Martin, 1972).

Since human sound patterns are inherently rhythmic, however, these notions cannot easily be tested using intact spoken or other humanly produced sounds. One possibility is to temporally distort connected speech sequences in some artificial way. Another is to use electromechanically timed nonspeech sounds, as w[illegible] in the present experiments. *Same–d*[illegible] judgments were used in the experimen[illegible] to get at the more perceptual aspects of input processing. However, inputs in memory experiments also are frequently sequential so that, in the present view, the results of memory experiments also can be affected by the temporal structure of the input. A conventional memory task—written recall—was used to show that this is the case.

The stimulus inputs used in the experiments were repeating auditory temporal patterns consisting of sequences of evenly spaced high- and low-buzz tones. Each pattern was either 7 or 8 elements in length, and began with a high tone and ended with a low tone. The first (high) tone was assumed to be heard as accented, as were all following run-initial high tones. Some patterns were rhythmic as defined by the rhythm rule presented elsewhere (Martin, 1972), and some were [illegible]t. For instance, a sequence of 8 equidistant elements is rhythmic for present purposes if accents are located on elements in Serial Positions 1 + 3 and/or 5, i.e., accents are 2 elements apart or some very simple multiple thereof. In consequence, hearing the early accents in a pattern generates expectancies concerning later accents. Additionally and just as important, these patterns generate expectancies concerning their termination (an even-numbered element), hence expectancies concerning when they begin repeating (an odd-numbered element). By contrast, an 8-element pattern is nonrhythmic if accents are located on Elements 1 + 4, since this pattern of accents generates initial expectancies for triplets.[3] Such expectancies will be violated, since triplets are appropriate for 9- (or 6-) element patterns but not for 8-element patterns. Further examples of nonrhythmic patterns are those having

[3] Obviously there cannot be accents on Positions 3 or 5 if there is one on Position 4, if accents are defined as run-initial high tones.

accents on Elements 1, 4, and 6, or on 1, 3, and 6. These patterns mix duples and triplets, hence generate either conflicting expectancies or else no expectancies in particular, including expectancies concerning the termination of the pattern. Thus, receiving no impression of a coherent pattern during listening, it should be more difficult for *S* to recognize the pattern as repeating or to recall it.

Besides rhythmic and nonrhythmic 8-element patterns, matching 7-element patterns were also used in the experiments. These are interesting because while they are shorter than 8-element patterns, hence on most theories should require less processing, they cannot have a simple rhythm, i.e., accents cannot be equidistant.[4] To demonstrate to yourself that 8-element patterns are simpler and more natural than 7-element patterns, tap an 8-element pattern on your desk, continuously and repeatedly, as fast as you can, accenting Taps 1, 3, and 5. Then try a 7-element pattern, again accenting Taps 1, 3, and 5. If not immediately convinced, practice a bit. For most people it may take some time before the 7-element patterns become smoothly and evenly repeated. On the present view, the relative naturalness of evenly spaced accents during production has consequences for perception also, which can be demonstrated under appropriate experimental conditions. For instance, a repeating 7-element pattern with accents on Elements 1 and 5, or on Elements 1, 3, and 5, generates expectancies during listening for an 8-element pattern which, however, will be violated since the expected pattern begins repeating one element too soon. In consequence, it should be harder to recognize the pattern as repeating. Additionally, with respect to memory, 7-element patterns should also be harder to recall than 8-element patterns, even if, as in these experiments, 7- and 8-element patterns are equivalent in number and location of accents, and in number of runs, "groups," "chunks," etc., dif-

[4] Of course, there are 7-beat musical rhythms, but these require more sophisticated responses in perception or production than do simpler rhythms.

TABLE 1

Experimental Patterns

Good		Poor	
8 element	7 element	8 element	7 element
10001110	1000110	10011100	1001100
11001000	1100100	11010000	1101000
11101000	1110100	11100100	1110010
11001110	1100110	11000110	1100010
10101000	1010100	10110110	1011010
10101110	1010110	11010100	1101010

Note. 1 = high tone and 0 = low tone. Good refers to rhythmic patterns and poor refers to nonrhythmic patterns.

fering only in that one group (or chunk) is smaller.

Experimental patterns were presented continuously, but only twice, i.e., repeated only once. The 7- and 8-element patterns were mixed. There were also dummy patterns that changed somewhat, i.e., did not repeat exactly, so that *S*'s first task was to judge *same–different*, i.e., repeated or changed. The second task was to write the pattern(s). The 3 experiments were identical except for presentation rate and response delay time, and will be discussed together whenever convenient.

Method

Experiment 1

Subjects and design. The *S*s were 44 undergraduates from the University of Maryland. Rhythmic structure (rhythmic vs. nonrhythmic) and pattern length (7 vs. 8 tones) were combined factorially as within-*S*s variables.

Materials. The buzz tones were approximately 180 Hz. (high) and 150 Hz. (low) presented at a 3.3/sec rate. The tones were 200 msec. long, with 100-msec. silent intervals between them.

The experimental, or repeating (*same*), patterns are shown in Table 1. Each begin with a high tone as preferred by most *S*s (Royer & Garner, 1970) and assumed to be heard as accented (Garner, 1970a). Looking at 8-element patterns first, the "good" (rhythmic) patterns had accents (single or run-initial high tones) on Serial Positions 1 and 5, and in 2 cases also on Position 3. The "poor" (nonrhythmic) patterns had accents on Positions 1, 4, and/or 6.

As Table 1 shows, each 8-element pattern was matched by a 7-element pattern constructed by deleting one item from the end of a run in its matching 8-element pattern at Serial Position 6 or 7. Thus, the 7- and 8-element patterns were exactly equivalent in the number of runs and in the number and location of accents. They were

TABLE 2

MEAN PROPORTION CORRECT JUDGMENTS

Experiment	Pattern			
	8-element-good	8-element-poor	7-element-good	7-element-poor
I (N = 44)	.80	.74	.77	.71
II (N = 24)	.80	.74	.73	.61
III (N = 42)	.80	.58	.69	.53

Note. Good refers to rhythmic patterns and poor refers to nonrhythmic patterns.

also equivalent in number and content of groups, chunks, etc., except that one group (or chunk) in the 7-element pattern was smaller (shorter). Each pattern was recorded twice continuously (repeated). Thus, each sequence was presented as a series of 14 equal-interval tones (the 7-element patterns) or 16 equal-interval tones (the 8-element patterns).

In addition to the 6 repeating (*same*) patterns of each type in the 2 × 2 design there were also 4 dummy patterns of each type that did not repeat but were changed slightly (*different*). These dummy sequences are not shown. For each of the dummy subsets, as in the experimental subsets (Table 1) there were 2 3-accent patterns and 4 2-accent patterns; in each of the subsets, half of the elements were high tones and half were low tones.

The 40 patterns (24 experimental and 16 dummy) were combined in a mixed list and were divided into 2 blocks of 20 patterns each with the following restrictions. Half of each subset of items was in each block and only one of each pair of matching 7- and 8-item patterns was in each block. In order to avoid repetition effects, whenever (half of) a dummy sequence was the same as an experimental pattern it always followed the experimental pattern in the block. Half of the *Ss* heard Block A first; the other half heard Block B first.

Procedure. The *Ss* first heard practice sequences with instructions to write the pattern. A nonrepeated pattern of each length was presented first, then 4 repeating patterns, one each of a 7- and 8-element pattern repeated (*same*) and changed (*different*). After hearing each sequence, *Ss* made their judgments, wrote the entire pattern, and were given feedback about the correct responses. No feedback was given during the presentation of the experimental patterns that followed

The *Ss* in small groups heard the sequences monaurally through headphones and responded after each on an answer sheet which for each sequence provided the letters *S* and *D* and 2 solid lines for writing the 2 patterns. The *Ss'* first task was to judge *same* or *different* and circle *S* or *D*. The *Ss'* second task was to write the entire sequence whether repeating or changed with 1 representing a high tone and 0 a low tone. They were told that each pattern began with a high tone and ended with a low tone, to guess if uncertain, and to mark an X when unable to guess.

Experiment II

The *Ss* were 24 undergraduates fr[illegible] versity of Maryland. This experiment [illegible] to Experiment I except that the tones [illegible] sented at the rate of 4.4/sec. The ton[illegible] 150 msec. long, with 75-msec. silent inte[illegible] between them.

Experiment III

The *Ss* were 42 undergraduates from the University of Maryland. This experiment was identical to Experiment II except that after each presentation of the sequence there was a 4-sec. unfilled forced-delay interval before *Ss* could respond. The answer sheet had 3 lines, labeled A, B, and C, for each sequence. Four seconds after the end of the sequence *E* held up a card showing either A, B, or C to indicate the line on which *S* was to write his responses.

RESULTS

The following results are for the experimental sequences only; data for dummy sequences are not reported.

Judgments

The mean proportions of correct judgments over the 2 trials are shown in Table 2. For all experiments there were more correct judgments for good patterns than for poor patterns, as predicted: Experiment I, F (1, 43) = 7.90, $p < .01$; Experiment II, F (1, 23) = 11.01, $p < .01$; and Experiment III, F (1, 41) = 56.38, $p < .001$. For Experiment I the judgments for the 8-element patterns were no[illegible] significantly better than for the 7-[illegible] patterns, F (1, 43) = 2.94, $p < .10$. For Experiments II and III this difference was significant: Experiment II, F (1, 23) = 11.20, $p < .01$, and Experiment III, F (1, 41) = 9.55, $p < .01$. There were no interactions. In short, good patterns were more easily recognized as repeating than were poor patterns in all 3 experiments. Except for Experiment I (slower presentation rate), 8-element patterns were more easily recognized as repeating than 7-element patterns, as predicted.

Pattern Recall

Two measures were analyzed. The first was the number of repeating patterns

TABLE 3

Mean Correct Pattern Recall

Experiment	Trials 1 and 2 combined				Trial 2 plus judgment correct			
	8-element-good	8-element-poor	7-element-good	7-element-poor	8-element-good	8-element-poor	7-element-good	7-element-poor
I	4.25	2.15	3.78	2.88	2.30	1.14	2.02	1.59
II	2.92	1.63	2.25	1.67	1.75	.96	1.00	.75
III	3.00	1.14	2.73	1.12	1.79	.69	1.43	.48

Note. Good refers to rhythmic pattern and poor refers to nonrhythmic pattern.

completely correct over the 2 trials combined, ignoring the second half of the written response and also whether or not the *same–different* judgment was correct. The Trials 1 and 2 combined data in Table 3 show that in all 3 experiments there was better pattern recall for the good than for the poor patterns: Experiment I, F (1, 43) = 29.27, $p < .001$; Experiment II, F (1, 23) = 11.51, $p < .01$; and Experiment III, F (1, 41) = 58.70, $p < .001$. Using this response measure, there was no significant overall difference between 8- and 7-element patterns on any of the experiments (although the direction was in favor of 8-element patterns in the last 2 experiments). In Experiments I and II there was a significant interaction: Experiment I, F (1, 43) = 8.71, $p < .01$, and Experiment II, F (1, 23) = 6.16, $p < .025$. These interactions mean that the longer patterns gain relatively more when the patterns have a good structure, namely, rhythmic.

The second written recall measure was the number of patterns completely correct on Trial 2 when S also had accurately judged the pattern as repeated. They have the advantage of practice and of more confident judgments and should be somewhat more reliable. The Trial 2 plus judgment correct data in Table 3 show that for all 3 experiments the good patterns were easier than the poor patterns: Experiment I, F (1, 43) = 16.59, $p < .001$; Experiment II, F (1, 23) = 6.51, $p < .05$; and Experiment III, F (1, 41) = 52.36, $p < .001$. In Experiment I there was an interaction, F (1, 43) = 7.76, $p < .01$, again showing the relative disadvantage of greater length when the patterns were nonrhythmic. For Experiments II and III the 8-element patterns were recalled significantly better than the 7-element patterns: Experiment II, F (1, 23) = 8.46, $p < .01$, and Experiment III, F (1, 41) = 7.43, $p < .01$.

In short, rhythmic patterns are generally easier to recall than nonrhythmic patterns. Extra length depresses recall for nonrhythmic (nonstructured) sequences but not for rhythmic sequences.

Discussion

Under the conditions of these experiments, auditory temporal patterns generally were more difficult to recognize as repeating when they were nonrhythmic than when they were rhythmic, and were often less easily recalled when nonrhythmic than when rhythmic. This result generally held even when the rhythmic patterns (8-element good) were made nonrhythmic, or at least rhythmically more complex, by simply reducing their length (7-element good). Considering the recognition (judgment) task first, these results appear to go against element-by-element registration as a useful perceptual strategy, either deliberate or "automatic," particularly when the input rate is fairly rapid (Experiments II and III). (This is not to say that the listener could never profit by attempting such a strategy.) It is also unlikely that the listener in any sense compares the 2 patterns in the sequence by somehow holding them as units in echoic memory. Also unlikely is that the differences between 7- and 8-element patterns could be explained by some measure of sequential complexity other than rhythmic complexity, since only on rhythmic grounds is it easy to see why taking one element out of a run (or subgroup) should make the pattern more complex. More likely is that the listener "feels" the coherence of a rhythmic pattern

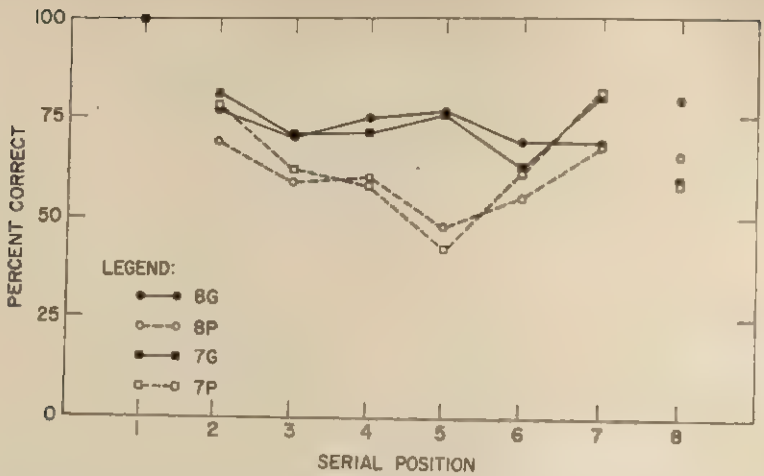


FIGURE 1. Serial position curves for the second trial of Experiment III. (The first 8 positions are shown, so that for 7-element patterns the last position represents the first element of the pattern as it was repeated. Abbreviations: G = good, i.e., rhythmic pattern, and P = poor, i.e., nonrhythmic pattern.)

as he hears it, and since this includes the "knowledge" that the pattern has terminated, hence the beginning of the next, the remainder of the sequence is more readily recognized as a repetition of the pattern just heard. This conclusion, that rhythmic patterns are somehow "easier" or more "natural," supports the hypothesis that listening mechanisms are optimally designed to handle sequential inputs that have a rhythmic organization.

It should be mentioned that an alternative experimental method would have been to present patterns of a given length in blocks, so that *S* would know in advance which length to expect, and hence could use various deliberate strategies. There is little doubt that *S* would do better with 7-element patterns than 8-element patterns under such conditions (though not better with poor patterns than with good patterns). This conclusion is hinted at in Table 1, which shows insignificant main effects of pattern length on judgments with the slower rate in Experiment I. But results favoring shorter patterns are not interesting in themselves since they are expected on nearly any theory of information processing.

Turning to written recall, consider the serial position curves in Figure 1, which were obtained from the second trial of Experiment III (similar curves were obtained from Experiments I and II). Position 8 was scored as correct for the 7-element patterns only when left blank. For 8-element patterns it was scored correct if 0 was recorded but incorrect either for 1 or a blank.

These curves are not quite conven[illegible]nal in several respects. For one, they repres[illegible] the first half of the sequences actually presented to *S* and his responses. For another the point representing the first tone is [illegible] connected to the following points, since [illegible] knew it was always high, and Position 8 in the sequence is not connected. The *S* would know whether this was high or low if he knew the length of the pattern he had just heard. Obviously he often did not, as can be seen by noting the poor scores at Positions 7 and 8.

Several aspects of these curves are of interest. For one, when Positions 1 and 8 are disregarded, the curves for good patterns are relatively flat. Poor patterns, on the other hand, show the dip in the middle typical of serial position curves for arbitrary sequ[illegible] of memory items. This result, however, may be in part a response bias toward writing good patterns, e.g., toward marking Position 5 as accented. Such a bias would result in an elevation at that position for good patterns and a depression for poor patterns, both tendencies of which can be seen in Figure 1.[5] Further suggestions for a response bias can

[5] This point was tested by linear contrasts of Position 5 against the average of Positions 4 and 6 for each of the 4 string types. All but the 8-element-good patterns were reliable ($p < .01$).

be seen in Table 3. Here, although the input stimuli were the same in Experiments II and III, the delay in Experiment III resulted in lower scores for 8- and 7-element-poor patterns, compared to Experiment II, but higher scores for 7-element-good patterns (8-element-good patterns remained unchanged). These results suggest a "drift" toward good patterns during the delay interval. Both of these tendencies, a flat serial position curve and a response bias in favor of good patterns, are to be expected on the view that rhythmic patterns are gestaltlike holistic entities.

Another result seen in Table 2 is worth mentioning. While the input sequences were exactly the same in Experiments II and III, the latter required *S* to make a delayed response. Comparing Experiments II and III, note that the delay depressed correct judgments for nonrhythmic patterns. Why should this happen, since presumably the judgments are made immediately even if the written response must wait? Most likely, the answer is that when patterns are nonrhythmic the impression of a coherent pattern is less firm and evaporates quickly over time, with the result that *S* now reports that the pattern did not repeat.

Finally, speculating on the implications of these experiments, the main point concerns the hypothesis that meaningful auditory patterns occurring in nature generally have a temporally redundant (rhythmic) organization, and that perceptual mechanisms have evolved to expect and make use of these redundancies. The importance of temporal structure can be seen in the fact that simple musical tunes remain recognizable when pitch variations are eliminated with timing relationships intact, but not vice versa (Lenneberg, 1967). A sequence (concatenation) of notes is not a tune unless the notes are in the right temporal places. Similarly, speech appears to be organized in such a way that the more informative linguistic elements are temporally predictable, i.e., organized around the accent structure of the sequence (Martin, 1972). It is natural to speculate that the temporal redundancies in speech at the level of sound facilitate relatively "automatic" processing at that level, thus freeing psychological resources for "attending" to the level of meaning. The fact that spontaneous speech departs from the received grammar, yet is understood nevertheless (Martin, 1967), is in line with this view. The listener can depend upon temporal, if not syntactic, regularities in the speech he hears.

REFERENCES

Garner, W. R. Good patterns have few alternatives. *American Scientist*, 1970, **58**, 34–42. (a)

Garner, W. R. The stimulus in information processing. *American Psychologist*, 1970, **25**, 350–358. (b)

Halle, M., & Stevens, K. N. Speech recognition: A model and a program for research. In J. A. Fodor & J. J. Katz (Eds.), *The structure of language: Readings in the philosophy of language*. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Lenneberg, E. H. *Biological foundations of language*. New York: Wiley, 1967.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431–461.

Martin, J. G. Hesitations in the speaker's production and listener's reproduction of utterances. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 903–909.

Martin, J. G. Rhythmic (hierarchical) vs. serial structure in speech and other behavior. *Psychological Review*, 1972, **6**, 487–509.

Miller, G. A. Decision units in the perception of speech. *I.R.E. Transactions on Information Theory*, 1962, **8**, 81–83.

Royer, F. L., & Garner, W. R. Perceptual organization of nine-element auditory temporal patterns. *Perception & Psychophysics*, 1970, **7**, 115–120.

(Received February 5, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 384–392

MEASURES OF INFORMATION PROCESSING IN CONCEPT IDENTIFICATION [1]

DALE E. BERGER [2]

Claremont Graduate School

A method was developed to show multiple-hypothesis behavior and states of information during concept identification (CI). Compared to Levine's 1966 blank-trial method, the new method provided more detail about information held by *S* at each stage of a problem and allowed examination of process assumptions of current CI models. Most *Ss* attempted to process information on all dimensions simultaneously; some initially selected a subset of dimensions to process. Processing errors were made following positive as well as negative feedback. No CI model could account for the data in detail, although protocols from individual *Ss* were consistent with aspects of the various models. A random-reversal assumption was suggested as a mechanism to account for departures from perfect processing.

Recent theoretical approaches have discussed behavior in concept identification (CI) tasks within the framework of hypothesis testing. Much of the work (e.g., Bower & Trabasso, 1964; Falmagne, 1972; Levine, 1966; Trabasso & Bower, 1968; Wickens & Millward, 1971) grew out of the "strategy selection" theory formulated by Restle (1962). This theory postulates sampling with replacement from a known set of hypotheses following each error.

Since the original investigation by Bruner, Goodnow, and Austin (1956), researchers have been concerned with the difficulties of externalizing strategies used in attaining concepts. This remains a most difficult problem. In a recent review of research on strategies in CI, Taplin and Jeeves (1972) concluded that valid and reliable indices of the processes of CI must be established if information-processing theories are to remain a promising approach to understanding concept attainment.

Levine (1966) introduced a technique that provides some information about hypothesis behavior of individual *Ss* i CI task. Assuming *S*'s responses betwe errors are determined by a single hypothesis, the pattern of responses to a set of blank trials (trials with no feedback) can be used to identify the hypothesis. Levine's blank-trial technique has been used extensively in testing implications of mathematical models of CI (cf. Erickson, 1968; Levine, 1969).

Levine's (1966) hypothesis-tracking technique will be called multiple-cue (MC) to contrast it with an alternative single-cue (SC) technique introduced in experiments to be reported here. The difference is in the complexity of the stimuli used for blank trials: The MC method uses complex stimuli which each have a value on every dimension, whereas the SC method uses stimuli which each have only one value on a single dimension (see Figure 1).

With the SC blank-trial technique, each stimulus dimension is represented in the blank-trial set by a single value presented in isolation. The *S* is given 3 response alternatives to SC stimuli: If he feels the cue is potentially relevant to solution, he is allowed to classify it into the correct category; if he has eliminated the dimension from the set of possible solutions, he would respond *irrelevant*; if he does not remember how the cue should be classified, he would respond *don't remember*. When a set of SC blank trials follows every feedback

[1] This article is based on a PhD dissertation submitted to the University of California, Los Angeles. A shorter version was presented at the meeting of the Western Psychological Association, Portland, April 1972. The research was facilitated in part by grants from the University of California, Los Angeles, Graduate Student Patent Fund and the Campus Computing Network to the author, and U.S. Public Health Service Grant MH-08741 to T. Trabasso. The author wishes to express his thanks to T. Trabasso, M. Friedman, and T. Wickens, who offered many valuable suggestions.

[2] Requests for reprints should be sent to Dale E. Berger, Department of Psychology, Claremont Graduate School, Claremont, California 91711.

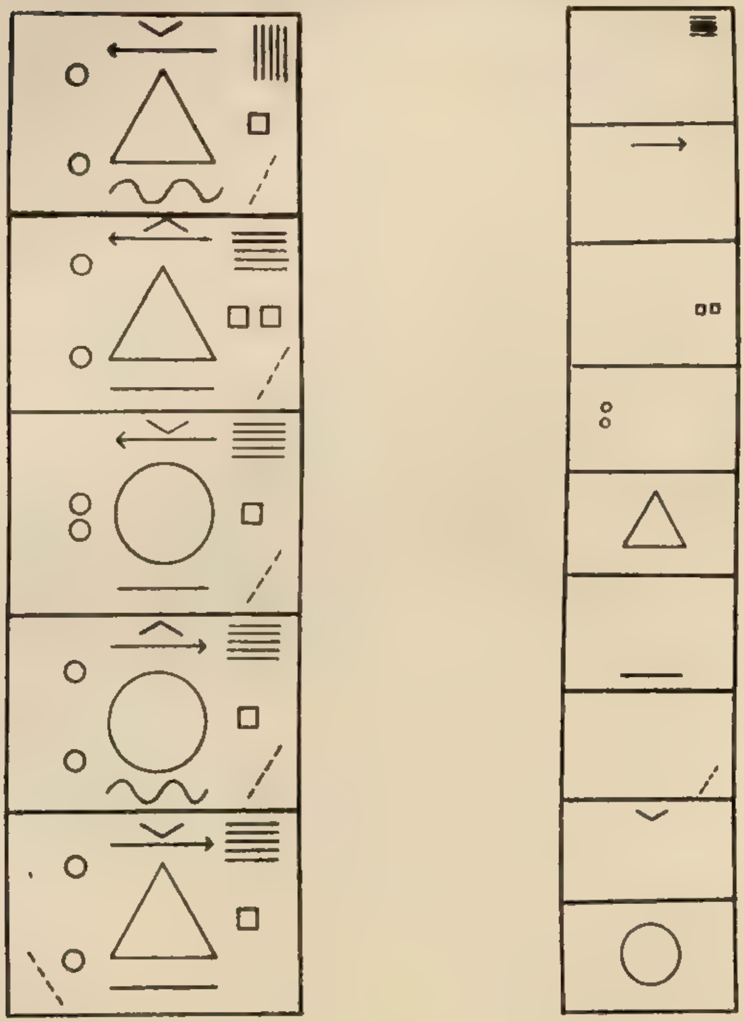


FIGURE 1. Examples of multiple-cue and single-cue blank-trial test sets.

trial (FT), it is possible to recover the information *S* has on each dimension at each stage of the problem.

The SC blank-trial procedure allows detailed examination of processing errors during the course of solving a CI problem. Eight types of errors account for all possible deviations from perfect processing. Table 1 shows the possible combinations of information held on a given dimension before an FT, the direction of the feedback, classification after the feedback, and the type of processing error, if any. If a dimension is thought to be relevant, it can be classified in either of 2 directions, labeled A and B in Table 1. Note that for the purpose of classifying errors, a *don't remember* response is here considered equivalent to *irrelevant* since both indicate that *S* is unable to classify the dimension.

The 8 types of errors or deviations from perfect processing in a simple binary CI task are as follows:

1. *Initial omission*—a failure to classify

TABLE 1
POSSIBLE ERRORS IN SIMPLE CONCEPT IDENTIFICATION PROBLEMS

Classification before feedback trial	Feedback (consistent with Category A or B)	Classification after feedback trial	Type of error (if any)
No hypothesis (before first feedback trial)	A	A	(Correct)
		B	Initial reversal
		Irrelevant	Initial omission
A	A	A	(Correct)
		B	Inconsistency reversal
		Irrelevant	Premature elimination
	B	A	Failure to eliminate
		B	Consistency reversal
		Irrelevant	(Correct)
Irrelevant	A	A	Consistent resample
		B	Inconsistent resample
		Irrelevant	(Correct)

a dimension in either direction following the first FT (FT_1), because information has either been forgotten or was never encoded.

2. *Initial reversal*—classification, following FT_1, of a value on a dimension opposite to the direction supported by FT_1.

3. *Failure to eliminate*—retaining a hypothesis following an FT where the hypothesis could logically be eliminated.

4. *Premature elimination*—elimination of a dimension before it could logically be eliminated.

5. *Consistency reversal*—reversal of the previous classification on a dimension, making the new classification consistent with the last FT. Since information from the last FT was inconsistent with information held previous to the FT, the dimension should have been eliminated.

6. *Inconsistency reversal*—reversal of the previous classification on a dimension, making the new hypothesis inconsistent with the last FT. Since information on the last FT was consistent with the previous hypothesis, this hypothesis should have been maintained.

7. *Consistent resample*—classification of a dimension which had previously been considered irrelevant to the solution in a way that is consistent with the last FT.

8. *Inconsistent resample*—classification of a dimension that had previously been considered irrelevant in a way that is inconsistent with the last FT.

Distributions of processing errors from individual *S*s can be used to test assumptions of various models of CI. For example, the hypothesis-manipulation model (Chumbley, 1969) assumes perfect processing following a correct response but massive resampling following a proportion $(1 - t)$ of error trials. Thus, resampling errors are the only errors predicted by the model, and they are expected to occur only following an error trial.

It should be noted that the MC blank-trial technique does not provide enough information to allow a comparable detailed examination of the assumptions of models. Only one hypothesis can be revealed by each MC blank-trial set, so it is impossible to determine which other hypotheses *S* may be tracking at that time.

The present experiment was designed to compare the MC and SC blank-trial techniques and to generate distributions of processing errors so models of CI could be evaluated. In a preliminary study with 4-dimensional CI problems, practiced college students performed so well that most made too few processing errors of any type to allow detailed analysis (Berger, 1970). However, performance was poorer on problems with interpolated MC blank trials than on problems with SC blank trials. The present experiment was an extension to more difficult problems with 8 dimensions. Differences between the SC and MC test procedures were reduced by signaling the trials on which *S* would receive feedback, and the effect of the interpolated tests on solution rate was evaluated by comparison with a control condition which had no interpolated tests.

METHOD

Subjects. The *S*s were volunteers from introductory psychology classes at the University of California, Los Angeles, who were fulfilling a class requirement. They were paid $1.50/hr for any time beyond their class requirement. Twenty-four *S*s participated in the experimental group, and 10 *S*s were used in a control condition.

Apparatus and materials. Stimuli were rear projected onto a screen in front of *S* by a Sawyer 550ER projector. A Davis timer was used to control the presentation of feedback. The response panel contained a button labeled "next" to advance the projector, and 4 labeled response choice buttons. There was 1 response button for each of 2 classification categories, a button labeled "irrelevant," and a button labeled "don't remember." The 2 category buttons had lights over them to indicate feedback. Also, there was a small light directly below the screen to signal the FTs.

The FT and MC test patterns each contained 8 binary attributes, and the SC stimuli each consisted of a single value (see Figure 1).

Problems. The stimulus sets used for FTs were arranged so that within each set every pattern differed from every other pattern on exactly half of the dimensions. Thus, an *S* who processed information perfectly would eliminate half of the remaining solutions on each FT. Since there are 16 elementary hypotheses in a problem with 8 binary dimensions, an *S* who eliminated half of the hypothesis following each FT would be left with a single hypothesis following FT_4. Five FTs were given in these experiments, since it was anticipated that most *Ss* would not solve these problems as rapidly as was logically possible. The problems were arranged so that each of the elementary hypotheses was the correct solution equally often.

Patterns in each test set for the MC blank trials met the requirements that (*a*) each stimulus differ from all other stimuli in the set on exactly half of the dimensions, and (*b*) both values of each dimension occur at least once in each test set. These restrictions insure that within each subset it is impossible for *S* to be correct every time if he responds on the basis of a response bias, and if *S* responds on the basis of a single hypothesis, his response protocol for any subset is unique for each of the possible hypotheses. Only 4 patterns are actually necessary in the subset to determine the hypothesis with which *S* responded, but following Levine (1966), a fifth pattern was included to provide a consistency check (see Figure 1).

The set of patterns for the SC test trials consisted of 9 stimuli, each composed of 1 value on a single dimension. The first 8 patterns included 1 value from each of the 8 dimensions, and the ninth pattern served as a consistency check. With equal frequencies, this last pattern was the complement of a value which had occurred earlier in Position 1, 3, 5, or 7 in the test set.

The problems alternated between those with SC and those with MC tests following each FT. The control condition had no interpolated tests, but both types of tests were given after the last FT in all conditions, and *S* was asked what he knew about the problem.

Procedure. Since the focus of the experiment was on asymptotic behavior, extensive training was given to *Ss*. Every *S* attempted 7 practice problems with 4 dimensions and 4 practice and 16 experimental problems with 8-dimensional patterns. The *Ss* in the control condition were given the same instructions and practice problems. Most *Ss* required less than 3½ hr.

TABLE 2

COMPARISON OF BEST AND POOREST *Ss* ON PROPORTION OF PROBLEMS SOLVED AND INCONSISTENCIES ON MULTIPLE-CUE (MC) AND SINGLE-CUE (SC) BLANK-TRIAL SETS

S	Solution rates[a]				Inconsistencies	
	1	2	3	4	MC	SC
Best[b]	.50	.66	.70	.70	.198	.041
Poorest[c]	.08	.06	.19	.23	.242	.173
Control[d]	.20	.38	.43	.45	—	—

[a] Blocks of 4 problems.
[b] $n = 11$
[c] $n = 13$
[d] $n = 10$.

After *S* responded to a pattern, he pressed a central button labeled "next." This stepped the projector and brought *S*'s hand to a neutral position. On FTs, this also caused a light to appear for 1 sec. between the label and button of the correct class. The feedback light appeared just after the stimulus was removed, and was off before the next stimulus was projected.

In response to an MC pattern, *S* had to press 1 of the 2 classification buttons. Given an SC pattern, *S* had 3 options. If he thought the value had always been consistent with only 1 category, he was to press the corresponding button. If the value had been paired with both categories, he was to press the *irrelevant* button. If he could not remember to which category the value belonged, he was to press the *don't remember* button.

RESULTS AND DISCUSSION

Each *S* attempted 16 experimental problems. After the last FT in each problem, every *S* was given an SC and MC test series and was asked what he knew about the problem. An *S* was considered to have solved a problem when the results from both cue tests agreed with the correct hypothesis. Individual differences were very large, with *Ss* solving 0–13 of the 16 problems. The 11 best *Ss* solved at least 7 problems, and none of the remaining 13 *Ss* solved more than 4 problems. Because of this natural break, *Ss* were divided into 2 groups of unequal numbers for subsequent analyses. Overall performance improved rapidly over the first half of the experiment, and was relatively stable during the last half (Table 2). By the

second block of 4 problems, the better *S*s were performing nearly as well as they did during the rest of the experiment. The poorer *S*s, however, did not show much improvement until the third block of problems, and it is likely that they were still improving at the end of the experiment.

Comparison of tests. The 2 types of interpolated tests yielded similar accounts of the composition of the hypothesis sets after FT_5. Of the 384 problems (16 problems from 24 *S*s), both tests agreed that the correct solution had been reached on 142 problems, and on 62 problems, both tests agreed that a particular incorrect hypothesis had been reached as a solution. On 149 other problems, the SC test indicated that *S* was tracking more than 1 hypothesis following FT_5. The remaining 31 problems showed inconsistencies on the MC test, failure to track any hypothesis on the SC test, or disagreement between the 2 tests.

Effect of interpolated tests. In contrast to the pilot studies where SC problems were easier than MC problems, the proportions of SC and MC problems solved were exactly the same, 71/192 = .37. The signal that *S*s in the present experiment received at the beginning of FTs to distinguish them from MC blank trials is probably a relevant factor here. On each trial, the stimulus was removed before feedback was given, so the signal probably helped *S* to avoid processing irrelevant information in the MC blank trials and thus made the 2 probe tests more comparable.

The effect of interpolated test series on overall rate of problem solving can be gauged by comparing the proportion of *S*s who solved each experimental problem with the control condition. Ten *S*s in the control condition solved the same problems as experimental *S*s, but the 5 FTs had no tests interpolated between them. There was no difference in the proportion of problems solved by control *S*s (58/160 = .36) and by experimental *S*s(.37), so the interpolated tests did not influence overall rate of solving.

Inconsistencies in multiple-cue tests. The overall proportion of MC test protocols that were not consistent with any single hypothesis was .222. There was no difference between the best and poorest *S*s, U (11, 13) = 56.5, $p > .25$ (see Table 2).

Inconsistencies in MC test protocols have been interpreted as being the result of random response errors, sometimes called "oops errors" (Levine, 1969). However, an analysis of the inconsistent response protocols showed that they were not all due to random response errors since many could be attributed to a response rule used by *S*s. The "response bias" rule of responding with the same category name to each stimulus in the MC test series was used more often than chance by 19 of the 22 *S*s who showed MC inconsistencies (one-tailed binomial, $p < .001$). Also, the likelihood of an inconsistency on the last MC test depended on the number of hypotheses that *S* was tracking. When the SC test following the last FT indicated that more than 1 hypothesis was considered to be possibly relevant, .161 of the MC test protocols were inconsistent, whereas only .041 were inconsistent when only 1 hypothesis was tracked on the SC test, $\chi^2(1) = 4.48$, $p < .001$.

Inconsistencies in single-cue tests. Nine blank trials were used in each SC test set. Each of the first 8 trials tested a different dimension, and the ninth provided a consistency check by retesting 1 of the 8 dimensions with the value opposite that which had occurred earlier. An inconsistency was said to have occurred if the test and retest did not give the same results. The best *S*s had a much lower proportion of SC inconsistencies over the 5 FTs than did the poorest *S*s, an average of .041 compared with .173 per test, U (11, 13) = 4, $p < .001$ (see Table 2). An SC inconsistency could indicate guessing, a response error, a hypothesis change, forgetting during the course of the set of blank trials, or that *S* had not integrated the information available on the given dimension so that it could be applied equally to both values.

It should be noted that an inconsistency does not necessarily have the same meaning

for an SC and MC test. An SC inconsistency means S failed to classify the 2 complementary values from a dimension in a consistent manner, while an MC inconsistency means S did not respond on the basis of a single hypothesis possibly because he used some other response rule.

Size of hypothesis set. The collection of hypotheses that S considers to be potentially relevant at any given time will be called his hypothesis set. This set does not include values from a dimension that has been ignored or forgotten. The SC tests reveal directly the contents of the hypothesis set at each point in the problem, and it is a simple matter to count the number of hypotheses. Figure 2 shows the mean number of cues in the hypothesis set for the best and poorest Ss following positive and negative feedback on each FT. On average, the poorest Ss started with fewer hypotheses in their sets, but were less successful in eliminating hypotheses than were the best Ss. The size of the hypothesis set was independent of the kind of feedback, in opposition to theories which assume that resampling is more likely to occur or will always occur following an error. It might be noted, however, that Ss in the present study were highly experienced.

Distributions of processing errors. An S who processes all information perfectly will start with 8 hypotheses after FT_1 and will eliminate all hypotheses which are not consistent with new information presented with each subsequent FT. As described in Table 1, there are 8 types of possible processing errors. Distributions of errors from each S as a function of positive or negative feedback on the preceding FT are shown in Table 3, with Ss arranged in descending order of success in solving the problems.

Overall, processing errors were more likely to follow negative feedback than positive feedback, $z = 4.06$, $p < .001$. This may appear to be consistent with many information-processing theories of CI which assume that most or all processing errors occur following negative feedback (cf. Chumbley, 1969; Trabasso & Bower, 1968). However, the large proportion of errors following positive feedback (.447) must be considered evidence counter to these theories.

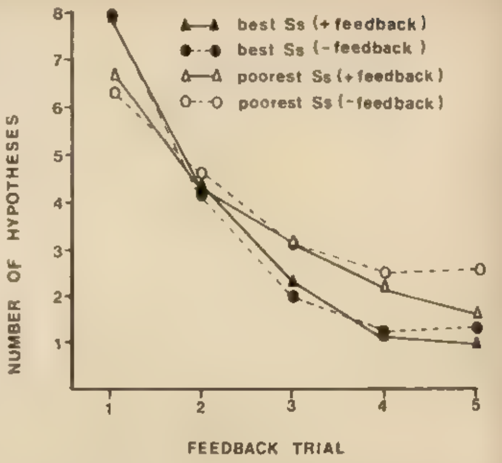


FIGURE 2. Size of the hypothesis set as a function of feedback trial number and feedback for the best and poorest Ss.

Strategies of individual Ss and departures from perfect processing can be identified from the distributions of errors. A particularly straightforward example is S_{24}, the last S in Table 3. His protocol clearly shows that he reduced his hypothesis set size to an average of 2 hypotheses at the beginning of the problems, and then used a strategy which required at least local consistency on resampling following errors. This S produced a total of 48 initial-omission errors on the 8 SC problems, for an average of 6 of these errors per problem. Since there were 8 dimensions in these problems and he eliminated an average of 6, S_{24} was tracking an average of only 2 of 8 dimensions following FT_1. This analysis was verified by a verbal report at the end of the experiment. When asked how he proceeded in his attempt to solve the problems, this S said he always started "with 1 or 2 parts." Most of the rest of the errors made by S_{24} were resampling cues which had previously been eliminated. These almost always followed only trials with negative feedback, as would be expected from an S who resamples only when the hypothesis he had been tracking was shown to be incorrect. The hypotheses he

TABLE 3

Distributions of Errors

Error type	Feedback	Ss[a] 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Totals
Initial omission	−	0	3	0	0	0	0	1	0	0	0	0	0	22	0	2	0	17	7	4	2	1	1	1	19	80
	+	0	0	0	0	0	0	0	1	0	0	0	0	5	0	8	0	2	4	1	6	0	3	2	29	61
Initial reversal	−	2	0	0	2	2	0	3	1	1	0	2	7	1	0	3	1	2	7	7	2	2	15	0	2	62
	+	0	0	0	0	7	1	0	0	1	0	1	3	12	0	0	1	3	2	6	5	3	4	1	0	50
Failure to eliminate	−	1	4	2	0	2	5	9	3	3	5	5	6	6	5	2	23	18	10	14	6	16	29	13	1	188
	+	2	0	2	0	0	6	8	7	1	2	3	3	2	14	2	7	12	15	12	1	22	7	3	1	132
Premature elimination	−	1	0	8	2	5	3	2	10	3	9	8	8	4	6	10	7	6	13	13	19	10	5	9	1	162
	+	3	0	3	0	6	2	8	5	5	5	3	14	7	7	11	2	9	12	14	7	19	10	10	1	163
Consistent reversal	−	2	0	0	1	1	2	4	9	4	1	0	3	3	5	0	2	9	4	8	3	3	5	11	0	80
	+	0	0	1	1	2	6	4	7	2	3	1	4	4	7	9	0	5	5	13	1	12	7	6	0	100
Inconsistent reversal	−	0	5	2	2	2	0	0	7	2	2	0	3	0	3	0	4	6	2	10	2	2	6	7	1	68
	+	0	0	1	0	0	2	1	1	1	0	1	1	0	2	1	4	0	1	4	4	5	4	2	1	36
Consistent resample	−	0	2	2	0	1	0	0	1	0	2	3	11	11	5	4	5	6	5	9	4	3	20	3	9	106
	+	0	0	1	0	2	0	2	1	4	4	3	1	9	6	7	2	9	5	4	6	2	2	7	2	79
Inconsistent resample	−	0	1	0	1	0	0	3	0	0	2	0	4	4	1	0	3	4	5	2	2	3	9	2	0	46
	+	1	0	1	0	0	0	1	1	0	1	0	1	1	0	0	1	4	2	1	1	1	3	0	0	20
Totals	−	6	15	14	8	13	10	22	31	13	21	18	42	51	25	21	45	68	53	67	40	40	90	46	33	792
	+	6	0	9	1	17	17	24	23	14	15	12	27	40	36	38	17	44	46	55	31	64	40	31	34	641

[a] Ss are arranged in descending order (1-24) of success in solving problems.

selected for his new hypothesis set were always consistent with the last feedback (no inconsistent-resampling errors). This is the "local consistency" strategy as described by Gregg and Simon (1967).

An *S* who used a selection strategy (cf. Restle, 1962) would select a subset of possible hypotheses following FT_1, producing many initial-omission errors. Only six *S*s had over 5 initial-omission errors, but these 6 *S*s produced 8–48 each (Table 3). Judging from the scarcity of initial-omission errors in the remaining protocols, the most popular initial strategy was a "global" strategy (cf. Bruner et al., 1956), which requires *S* to process information on all dimensions simultaneously. Although the best *S*s used this strategy with considerable success, the poorer *S*s were overextended and generally made many premature-elimination and failure-to-eliminate errors.

It is apparent, however, that no *S* performed perfectly in accord with any current CI model. Failure to predict error protocols in detail is not surprising in that models typically do not include mechanisms to account for departures from idealized protocols. Such a mechanism is suggested by the current data and could be included in any model.

The random-reversal assumption. Information held by *S*s attempting to solve a CI problem may be considered in 3 parts: classification of a given dimension before an FT, cue value and feedback received from the FT, and classification of the dimension following the FT. The random-reversal assumption states that there is a small probability that information at any of these 3 steps may be lost or reversed. An alternative, suggested by Falmagne's (1972) model, is that *S* has a confidence hierarchy for the information he holds on the various dimensions, and the probability of information loss or reversal on a dimension is inversely related to the confidence.

The consequences of information loss or reversal at any 1 of the 3 steps may be examined with reference to Table 1. Reversal of the classification before an FT can lead to 1 of 2 types of errors, depending on subsequent feedback: If the feedback is consistent with the new (reversed) classification, the error will be consistency reversal; if the feedback is inconsistent with the reversed classification, the error will be premature elimination. Reversal of the classification following FT_1 leads to an initial-reversal error; reversal following a subsequent FT leads to an inconsistency-reversal error.

Perhaps the most interesting predictions of the random-reversal assumption concern the consequences of reversal of information presented on an FT. If the information is changed to be consistent with the previous classification, the error will be failure to eliminate; if the information is changed to be inconsistent with the previous classification, the error will be premature elimination. These 2 types of errors were by far the most common; together, they accounted for almost half of the total errors (Table 3).

Reversal of information could be due to incorrect encoding of the cue value or of the feedback direction; it could indicate a logical error in combining new information with old, or a memory failure followed by an incorrect guess. Any of these errors would contribute to degradation of ideal performance predicted by a simple model.

An initial-omission error is probably less an error than the result of a decision to process fewer than the total number of available dimensions. Since *S*s differ widely on this strategy (Table 3), it may be necessary to include initial sample set size as a parameter in CI models.

The SC procedure on highly trained individual *S*s was successful in providing direct evidence on a number of processing errors that have relevance to the kind of information-processing strategy that *S*s use in selecting and testing hypotheses in inductive reasoning such as CI. The present analysis has shown that individual *S*s might exhibit any of several alternative strategies, each of which is formalized in particular assumptions of existing models. In a sense, no model is entirely correct or wrong, since some protocols favor or contradict each model in turn. What appears to be lacking in the models are mechanisms such as forgetting or reversing information

so that the idealized or formal strategies are not consistently exhibited. The present experiments and specific information-processing model indicated what erroneous behavior occurs and which mechanisms seem to be responsible for the departures from perfect or consistent processing.

REFERENCES

Berger, D. E. *Direct measures of information processing in concept identification.* (Doctoral dissertation, University of California, Los Angeles) Ann Arbor, Mich.: University Microfilms, 1970, no. 70–19, 828.

Bower, G. H., & Trabasso, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford University Press, 1964.

Bruner, J. S., Goodnow, J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Chumbley, J. I. Hypothesis memory in concept learning. *Journal of Mathematical Psychology,* 1969, **6**, 528–540.

Erickson, J. R. Hypothesis sampling in concept identification. *Journal of Experimental Psychology,* 1968, **76**, 12–18.

Falmagne, R. Memory process in concept ide fication. *Journal of Experimental Psychol* 1972, **92**, 33–42.

Gregg, L. W., & Simon, H. A. Process models a stochastic theories of simple concept formatio *Journal of Mathematical Psychology,* 1967, 4 246–276.

Levine, M. Hypothesis behavior by humans durin discrimination learning. *Journal of Experiment Psychology,* 1966, **71**, 331–338.

Levine, M. Neo-noncontinuity theory. In G Bower & J. T. Spence (Eds.), *The psychology o learning and motivation.* Vol. 3. New York Academic Press, 1969.

Restle, F. The selection of strategies in cue learning. *Psychological Review,* 1962, **69**, 329–343.

Taplin, J. E., & Jeeves, M. A. *Strategies in concept learning.* (Program on Concept Learning Rep 19) Boulder: University of Colorado, Institut for the Study of Intellectual Behavior, 1972.

Trabasso, T., & Bower, G. H. *Attention in learning: Theory and research.* New York: Wiley, 1968.

Wickens, T. D., & Millward, R. B. Attribute elimination strategies for concept identification with practiced subjects. *Journal of Mathematical Psychology,* 1971, **8**, 453–480.

(Received December 9, 1972)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 393–398

ABSTRACT CONCEPT LEARNING IN THE PIGEON [1]

THOMAS ZENTALL [2] AND DAVID HOGAN

University of Pittsburgh

In 1961, Cumming and Berryman found that color matching-to-sample training with pigeons did not transfer to a new color. In Experiment I in the present study, pigeons were trained on either a 2-color matching-to-sample task or a 2-color oddity-from-sample task. In the transfer phase, the 2 training colors were replaced by 2 novel colors for all *S*s and half of each training group was shifted to the other task. Nonshifted *S*s learned the transfer tasks faster than shifted *S*s. In Experiment II, using the same design, similar results were obtained when *S*s were trained on either brightness matching or brightness oddity and transferred to color matching or color oddity. Apparently pigeons are capable of learning the concepts *same* and/or *different*.

In a now classic experiment Herrnstein and Loveland (1964) demonstrated that pigeons could learn to discriminate between the presence of a human form (positive stimulus) and its absence (negative stimulus) thus suggesting that pigeons are capable of learning complex concepts. Since a large number of quite different complex visual images (photographs of natural settings) were used in both the positive and negative set, the task could not have been mastered by learning a response to each of the individual stimuli. It is less clear, however, that there was not some common attribute, other than the concept "person-present," which served to distinguish the positive from the negative set. While there was no attempt to equate the positive and negative sets in all other respects, the authors were aware of this problem. For example, to show that color distribution was not the basis for the discrimination, they presented black-and-white photographs, with little disruption in discrimination. Siegel and Honig (1970) and Poole and Lander (1971) extended these findings and ruled out a number of other artifactual cues upon which the discrimination might have been made. But, given the complexity of the stimuli used in all 3 experiments, the adequate test for, or control of, all factors other than person-present, may be virtually impossible.

A more efficient means of measuring concept learning in pigeons is with the use of simple stimuli which vary along a single physical dimension and a concept which involves the relation among these stimuli. The matching-to-sample task potentially satisfies these conditions.

In a matching-to-sample task an organism is presented with a sample stimulus and 2 or more comparison stimuli. The task requires the selection of that comparison stimulus which is the same as (matches) the sample stimulus.

Cumming and Berryman (1961) trained pigeons on a 3-color matching-to-sample task (red, green, and blue). Responses to the sample stimulus which was presented on the center response key produced the 2 comparison stimuli, one on each side key. A response to the comparison stimulus which matched the sample was reinforced.

At least 3 different means of learning this task can be described. (*a*) Pigeons may learn to make a particular response to each different stimulus configuration (e.g., peck the left key when presented

[1] This research was supported in part by Biomedical Sciences Support Grant FR7084 from the National Institutes of Health to the University of Pittsburgh and by National Institute of Mental Health Grant MH19757 to the first author. Portions of Experiment I were reported at the meeting of the Psychonomic Society, St. Louis, 1971. The authors are grateful to Celeste Compomizzi, Karen Compomizzi, Patricia Hughes, and Nancy Threlfall for their assistance with experimental sessions.

[2] Requests for reprints should be sent to Thomas Zentall, Department of Psychology, University of Pittsburgh, Pennsylvania 15260.

with the stimulus configuration, red–left, red–center, green–right). With 3 sample stimuli the pigeons would learn to respond individually to the 12 different stimulus configurations. (*b*) Pigeons may learn to form a response chain, one for each of the 3 sample stimuli (e.g., peck red on the center then peck red on the side). (*c*) Pigeons may learn a matching concept (i.e., peck that comparison stimulus which is the same as the center stimulus).

In an effort to determine what was learned during matching training, Cumming and Berryman (1961) substituted yellow stimuli for the blue stimuli. They found that performance was disrupted (dropped to chance) only on trials involving a yellow sample stimulus, thus suggesting that the pigeons learned response chains. The pigeons responded appropriately to red and green sample stimuli but, since they had not learned a response chain for yellow, they performed at chance when yellow was the sample. Cumming and Berryman concluded that matching training did not result in the formation of a matching concept. However, an alternative interpretation can be stated. If in fact there were competing response tendencies on yellow-sample trials, (*a*) to respond to that stimulus which had been reinforced in the past (i.e., the nonmatching red or green stimulus), and (*b*) to respond to the stimulus which matched the sample stimulus (i.e., the matching yellow stimulus), then performance may have reverted to chance, either because the pigeons alternated between the 2 response tendencies or, more likely, because the pigeons, under conflict, adopted position-response tendencies.

The purpose of the present experiments was to determine whether pigeons can acquire a matching (and/or oddity) concept if transfer involved only novel stimuli. Experiment I examined the extent to which color matching and oddity would transfer to novel colors. Since such transfer might be explained by simple stimulus generalization, rather than complex concept learning, Experiment II measured the extent to which matching and oddity training would transfer to stimuli not falling on the dimension used during training. To control for nonspecific transfer effects half of the *Ss* were initially trained on an oddity-from-sample task and shifted to the matching task with the novel stimuli. In the oddity task a response to the comparison stimulus which is *different* from the sample stimulus, is reinforced. Thus, learning an oddity concept should *retard* learning a matching task with new stimuli, and learning a matching concept should *facilitate* learning a matching task with new stimuli. Similarly, learning an oddity concept should facilitate learning an oddity task with new stimuli, and learning a matching task should retard learning an oddity task with new stimuli. Experiment I involved 4 groups of *Ss*. Two groups were trained on a 2-color matching task, and one of these was transferred to another matching task, while the second was transferred to an oddity task. Similarly, 2 groups were trained on a 2-color oddity task; 1 group was transferred to another oddity task and the second was transferred to a matching task. The transfer task involved 2 novel colors and transfer sessions were continued to assess learning.

Experiment I

Method

Subjects. Eight experimentally naive female white Carneaux pigeons approximately 1 yr. old, were maintained at 80% of their free-feeding weights throughout the experiment.

Apparatus. A standard 3-key Lehigh Valley Electronics Pigeon Test Chamber (1519D) was used. In-line projectors behind each key (with G. E. No. 1820 lamps) projected red (Kodak Wratten filter No. 26), green (Kodak Wratten filter No. 60), blue (Kodak Wratten filter No. 38A), or yellow (Kodak Wratten filter No. 9) fields onto each of the keys. Control equipment was located in the next room. Reinforcement consisted of 3-sec. access to Purina Pigeon Grains.

Procedure. All *Ss* were magazine trained and shaped to peck a lighted side key on the first day. Shaping position (left or right) and color (red or green) were counterbalanced over *Ss*. On the second day, all *Ss* were given 48 continuous reinforcements for responding to a single lighted key which varied in color (red or green) and/or position (left or right) following each reinforcement. On the third day, half of the *Ss* received matching

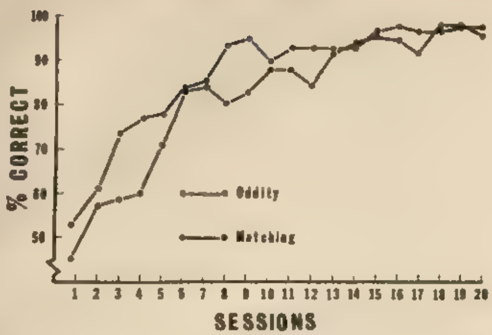


FIGURE 1. Comparison of matching and oddity performance during training: Experiment I.

training while the other half received oddity training. Each trial started with the onset of the sample stimulus on the center key. Five responses to the center key turned on the side keys. A single response to either side key terminated the trial and initiated a 5-sec. intertrial interval. Stimulus presentations for the 2 tasks were the same, only the "correct" response differed. Daily sessions consisted of 96 trials, counterbalanced for correct color and correct side. Twenty such training sessions were given.

Transfer sessions began following the last day of training. All *S*s were switched from red and green stimuli to yellow and blue stimuli. Half of the *S*s continued with the same task as during training (matching or oddity) while the remaining *S*s were shifted to the other task. Eleven transfer sessions were run

Results

Training performance data are presented in Figure 1. The data suggest that the oddity task was learned faster than the matching task, though this difference was significant only on the first session, $F(1, 6) = 6.40$, $p < .05$. On the first session, 3 of the 4 oddity *S*s performed above chance while 3 of the 4 matching *S*s performed *below* chance. By the end of training no difference in performance between the 2 groups was apparent.

The transfer data are presented in Figure 2. The nonshifted *S*s were those for which the training task and transfer task were the same; for the shifted *S*s the training task and transfer task differed. A 2-way analysis of performance on the first transfer session indicated a nonsignificant training effect (matching vs. oddity) ($F < 1$), a nonsignificant transfer effect (matching vs. oddity) ($F < 1$), and a nonsignificant Training × Transfer interaction (shifted vs. nonshifted), $F(1, 4) = 2.41$, $p > .05$.

The nonsignificant difference between shifted and nonshifted *S*s on the first transfer session may be due to the fact that 4 of the 8 *S*s were correct on exactly half of the trials. For the 4 birds whose performance differed from chance, however (1 shifted *S* and 3 nonshifted *S*s), performance of the shifted *S* was below that of the 3 nonshifted *S*s. A second 2-way analysis of performance pooled over the first 5 transfer sessions indicated a nonsignificant training effect ($F < 1$) and a nonsignificant transfer effect, $F(1, 4) = 1.80$, $p > .05$, but a significant Training × Transfer interaction, $F(1, 4) = 7.89$, $p < .05$. The interaction indicates that nonshifted *S*s performed significantly better over the first 5 transfer sessions than did the shifted *S*s.

Experiment II

Using the same design as in Experiment I, *S*s were trained on either a matching or oddity task with 4 sample brightnesses. Four brightnesses were used to

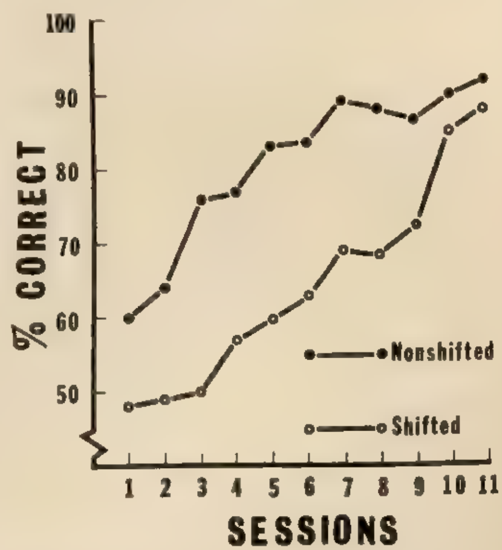


FIGURE 2. Experiment I transfer data. (Shifted *S*s are those transferred to a new concept—matching to oddity or oddity to matching; nonshifted *S*s are those maintained on the same concept—matching to matching or oddity to oddity.)

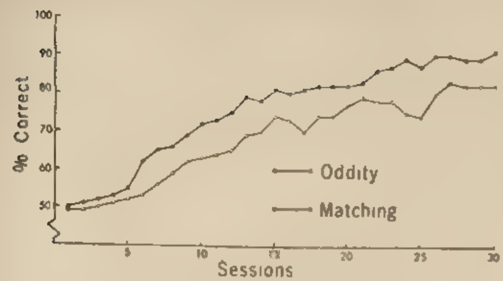


FIGURE 3. Comparison of matching and oddity performance during training: Experiment II.

maximize the likelihood that a matching and/or oddity concept would be formed. All *S*s were then transferred to 2 colors with half of the *S*s shifted in concept and the other half nonshifted.

Method

Subjects. Twelve experimentally naive female loft-reared pigeons approximately 1 yr. old were maintained at 80% of their free-feeding weights throughout the experiment.

Apparatus. The apparatus was the same as that used in Experiment I. The training stimuli were obtained by lighting 10 unfiltered lamps (G. E. No. 1820, supplied by 24 v.), 1 unfiltered lamp, 1 lamp with a 1.0 neutral density Kodak Wratten filter, or 1 lamp with a 2.0 neutral density Kodak Wratten filter. The brightness values were 13.0, 3.00, .30, and .03 ftc. as measured by a United Detector Technology Opto-meter (Model 40A) with a photome ric filter having transmission characteristics approximating the spectral sensitivity of the human eye. It should be noted that spectral sensitivity in pigeons is quite similar to that in humans (Blough, 1957).

The transfer stimuli were the red and green values used in Experiment I equated for brightness at 1.0 ftc. All brightness values were checked at the end of the experiment and were found to be consistent with initial measurement.

Procedure. On the first day, each *S* was shaped to peck either the right or left response key illuminated with 1 of the 4 brightness values. Shaping stimuli were assigned to *S*s such that, when *S*s were divided into 2 groups, 3 from each group had been shaped to the left and 3 to the right. Within each of the 2 groups 1 *S* had been shaped to the brightest stimulus value, 1 *S* to the dimmest stimulus value, and 2 *S*s each to the 2 middle stimulus values.

On the second day, all *S*s were given 48 continuous reinforcements for responding to a single lighted key which varied in brightness and/or position (left or right) following each reinforcement. On the third day, half of the *S*s received matching training while the other half received oddity training. With 4 sample brightness values and [...]dd brightness values (appearing either on th[...] eft or right) for each sample, 24 different sti[...] lus configurations could be presented. Each con[...]ration was presented 4 times in each 96-trial se[...]on. In all other respects training was the same [...] in Experiment I. All *S*s received 30 training se[...]ns.

Transfer sessions began following the last [...] of training. All *S*s were transferred from brig[...]ess stimuli to red and green stimuli matche[...] for brightness. Half of the *S*s continued with [...]eir training task (matching or oddity) while t[...] remaining *S*s were shifted to the other task. [...] *S*s received 6 transfer sessions.

Results

Training data for the matching an[...] dity tasks are presented in Figu[...] There is little evidence for matc[...] oddity differences on Session 1 ([...] but over sessions matching perform[...] was superior to oddity, F (1, 10) = [...] $p < .05$.

Transfer data are presented in Figur[...] A 2-way analysis of the data from [...] first transfer session indicated a significa[...] transfer effect, F (1, 8) = 13.78, $p <$.[...] and a significant Training × Transfer i[...]teraction, F (1, 8) = 19.26, $p < .01$, b[...] a nonsignificant training effect ($F <$ 1[...] The transfer effect indicates that initial performance on the oddity task was su[...]

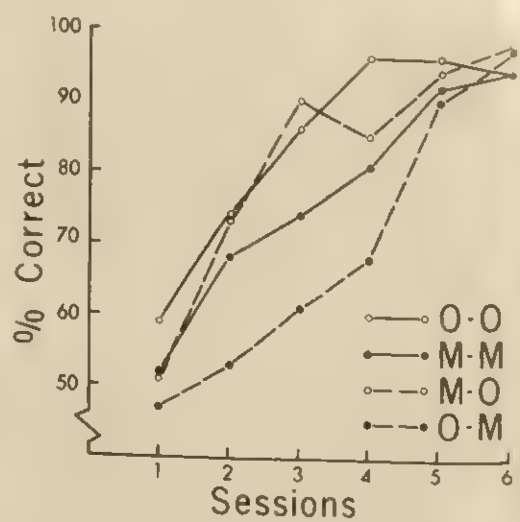


FIGURE 4. Experiment II transfer data. (Abbreviations: M-M = matching to matching, M-O = matching to oddity, O-M = oddity to matching, and O-O = oddity to oddity.)

perior to performance on the matching task. The Training × Transfer interaction indicates better performance for nonshifted *S*s than for shifted *S*s. A similar analysis of the Session 2 data indicated nonsignificant effects of training ($F < 1$), transfer, $F\ (1, 8) = 1.98$, and Training × Transfer interaction ($F < 1$). By the second transfer session, all but 1 *S* were performing above chance, and by Session 6 mean performance was above 95% correct. A 2-way analysis of the transfer data pooled over days indicated nonsignificant effects of training ($F < 1$), transfer, $F\ (1, 8) = 3.67$, $p > .05$, and Training × Transfer interaction ($F < 1$).

A 2-way analysis using sessions-to-criterion (90% correct) indicated similar results, though the transfer effect was significant, $F\ (1, 8) = 5.60$, $p < .05$, with this measure. Thus, for the transfer task, oddity was apparently learned faster than matching.

Discussion

The results of Experiment I indicate that matching and oddity training with pigeons will transfer better to new stimuli when the transfer task is the same as the training task than when the transfer and training tasks are different. The results of Experiment II indicate that these effects can be observed under conditions where generalization along a stimulus continuum (e.g., wavelength) cannot account for the transfer difference. The data suggest that the concepts *same* and/or *different* can be learned when pigeons master matching and oddity tasks, respectively. While the present design does not allow for the evaluation of these concepts separately, it does offer presumptive evidence for some form of "abstract" concept learning in pigeons. The word *abstract* is used here to identify transfer which cannot be explained by (*a*) stimulus generalization along physically definable dimensions or (*b*) nonspecific transfer effects.

While Experiments I and II are in basic agreement with regard to the transfer of concept learning, certain differences in the obtained results should be discussed.

During original training in Experiment I oddity performance was superior to matching performance for the first training session. In Experiment II there was vitually no difference between matching and oddity performance on the first training session, though over sessions superior performance was found on the matching task. These results are not inconsistent with the results of other studies which have reported comparisons of matching and oddity training.

Using a very lax criterion for learning (16 out of 20 correct), Ginsburg (1957) found that pigeons learned an oddity task faster than a matching task. Berryman, Cumming, Cohen, and Johnson (1965) found that performance on an oddity task was initially superior to performance on a matching task, but after about 600 training trials a crossover occurred and matching performance continued to be superior from then on. Both studies used 3 different sample stimuli for both matching and oddity. While oddity performance may be initially better than matching, the data suggest that matching may be learned faster when a moderate to stringent learning criterion is used.

Initial superiority of birds trained on the oddity task (Experiment I in this study; see, also, Berryman et al., 1965; Ginsburg, 1957) may be due to an artifact of pretraining. Typically, pretraining consists of single-stimulus presentation of each of the 3 keys (only 2 keys in the present studies). Then, at the start of training the center key is illuminated and responses to it illuminate the side keys, but for the first time these responses are not reinforced. It was observed in the present studies that during early training trials as many as 100 center key pecks occurred before a response was made to either side key (which terminated the trial). Undoubtedly, these nonreinforced pecks served to partially extinguish responding to the center key, and it is reasonable to expect that this partial extinction generalized to the same stimulus on the side key. Thus, better than chance performance on the oddity task can occur without any specific oddity training.

During training in Experiment II the failure to observe initial superiority of oddity over matching may be due to the specific training stimuli used. It may be that the brightness differences in Experiment II were not as salient as the color differences in Experiment I, and thus, initial extinction to the stimulus on the center key generalized nearly equally to the 2 comparison stimuli in Experiment II.

The superiority of matching later in training suggests that in addition to whatever concept

learning occurs, pigeons master matching and oddity by learning response chains (e.g., peck red on the center then peck red on the side). The fewer response chains to be learned the easier the task. In general, with n different sample stimuli, matching requires learning n response chains (one for each sample stimulus). Oddity, however, requires learning $n(n-1)$ response chains (a separate chain for each sample stimulus combined with each possible odd stimulus). With 2 sample stimuli, $n = 2$ and $n = n\ (n-1)$, matching and oddity should not differ in difficulty. Experiment I of the present study confirms this prediction. With 3 or more stimuli, $n < n\ (n-1)$ and matching should be easier than oddity. Berryman et al. (1965) and Experiment II of the present study confirm this prediction. Unfortunately the above model is insufficient to account for superior oddity learning (sessions to criterion) during transfer sessions in Experiment II. Other factors apparently contribute to the relative speeds of matching and oddity learning.

A second difference between the results of Experiments I and II was the occasion of the appearance of the concept–transfer effect. In Experiment I a significant transfer effect appeared only after the first transfer session; in Experiment II the same effect was significant only *on* the first transfer session. The difference in transfer results between the 2 experiments was probably due to the difference in transfer stimuli (yellow and blue in Experiment I; red and green in Experiment II), though it is possible that differences in either *S* pool (strain differences) or differences in original training, i.e., training stimuli (colors in Experiment I and brightness values in Experiment II), number of training stimuli (2 in Experiment I and 4 in Experiment II), and number of training sessions (20 in Experiment I and 30 in Experiment II) may have affected the transfer results as well. The fact that 2 *S*s in Experiment I responded to the yellow comparison stimulus on every trial of the first transfer session, while no *S* in Experiment II showed a complete color preference, suggests that differences in transfer stimuli may have been responsible for the observed differences in transfer results.

REFERENCES

Berryman, R., Cumming, W. W., Cohen, L. R., & Johnson, D. F. Acquisition and transfer of simultaneous oddity. *Psychological Reports*, 1965, **17**, 767–775.

Blough, D. S. Spectral sensitivity in the pigeon. *Journal of the Optical Society of America*, 1957, **47**, 827–833.

Cumming, W. W., & Berryman, R. Some data on matching behavior in the pigeon. *Journal of the Experimental Analysis of Behavior*, 1961, **4**, 281–284.

Ginsburg, N. Matching in pigeons. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 261–263.

Herrnstein, R. J., & Loveland, D. H. Complex visual concept in the pigeon. *Science*, 1964, **146**, 549–551.

Poole, J., & Lander, D. G. The pigeon's concept of pigeon. *Psychonomic Science*, 1971, **25**, 157–158.

Siegel, R. K., & Honig, W. K. Pigeon concept formation: Successive and simultaneous acquisition. *Journal of the Experimental Analysis of Behavior*, 1970, **13**, 385–390.

(Received February 22, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 399-408

DYNAMIC PROCESSES IN STIMULUS INTEGRATION THEORY: EFFECTS OF FEEDBACK ON AVERAGING OF MOTOR MOVEMENTS[1]

KENT L. NORMAN[2]

University of California, San Diego

A stimulus integration experiment was conducted which provided feedback to *S* in a task requiring the averaging of 2 fixed-movement lengths. Feedback indicated the relative position of each response to a weighted average of the lengths. Changes in the integration rule as a function of feedback were indicated by changes in the weighting parameters of a general weighted average model. Feedback appropriate to the numerical average reduced an initial bias toward the longer distance, while no changes were noted for a group that received no feedback. When feedback placed greater weight on the longer or shorter movement in a pair, systematic changes were observed in *S*s' responses, but not when feedback placed greater weight on the movement in the first or second ordinal position. These results indicate that in some cases the integration rule may be a function of dynamic processes such as feedback.

A judgment may be viewed as a response dependent on a set of stimuli, and the observer, an integrator of stimulus information. Stimulus integration theory and functional measurement proposed by Anderson (1962, 1970) has studied integration behavior in a number of tasks. The primary assumption of this theory is that the judgment is an algebraic function of the subjective stimulus values, s, and weights, w, on the subjective stimulus values.

While the observer may be viewed as an integrator of information, it must also be remembered that he is an organism capable of learning, and that the judgment response may also be a function of practice and feedback. It is likely that such feedback affects the integration function. Slovic and Lichtenstein (1972) pointed out that little or no research has been done on learning to integrate stimuli and on the effect of feedback in stimulus integration paradigms. However, experimentation on dynamic processes has been conducted in Bayesian decision-making and multiple-cue learning tasks. Methodologically, integration theory has an advantage over correlational paradigms in that it uses subjective scale values rather than physical values. Other advantages have been discussed elsewhere (Anderson, 1962, 1972).

The present research is an attempt to extend the investigation of the effects of feedback and dynamic processes to an integration task involving the averaging of motor movements. The *S*s were blindfolded and required to make several successive controlled motor movements and then make an additional linear movement to represent the average of the previous fixed movements. This task has been studied by Levin, Craft, and Norman (1971) and by Levin, Norman, and Dolezal (1973). These authors noted that a possible advantage that this task may have over other integration tasks is that stimulus magnitude is manipulated along a continuous dimension and responses are executed along the same dimension.

Levin et al. (1971) found that the data could be approximated by an ordinal-

[1] This project was supported during the author's tenure in the U.S. Public Health Service predoctoral fellowship, National Institute of Mental Health, under Grant MH-51134-01, and completed while a visitor at the Center for Human Information Processing, University of California, San Diego, under Grant MH-15828 from the National Institute of Mental Health. The report is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Iowa. Appreciation is expressed to the chairman, Irwin P. Levin, and members of the dissertation committee, Donald D. Dorfman, James V. Hinrichs, Raymond H. Hohle, and Rudolph W. Schulz, for their many helpful comments.

[2] Requests for reprints should be sent to Kent L. Norman, who is now at the Department of Psychology, University of Alabama, University, Alabama 35486.

dependent weighted average model,

$$R_A = \sum w_i s_i, \qquad [1]$$

where R_A is the averaging response, w_i is the weight varying with ordinal position on the subjective value s_i, such that $\sum w_i = 1$. However, there was a significant interaction between stimulus levels that could not be accounted for with an averaging model with constant weights. Levin et al. (1973) confirmed that the interaction between stimuli within a set was highly reliable. Since the response scale appeared reliable, a different averaging model was proposed that weighted stimuli differentially as a function of their value. This model can be expressed as

$$R_{ij} = \frac{v_i s_i + v_j s_j}{v_i + v_j}, \qquad [2]$$

where R_{ij} is the magnitude of the averaging response for stimuli i and j, and where v_i and v_j refer to weights on the subjective stimulus values varying with the values of the stimuli. Furthermore, the restriction is added that the weights sum to 1. The fact that both order-dependent weighting and value-dependent weighting may be present in this motor task make it an ideal one to study.

The following 3 issues relating to the effect of feedback on the integration rule were investigated in the present study: (*a*) effects of "correct" numerical feedback vs. no feedback, (*b*) effects of feedback on order-dependent weighting, and (*c*) effects of feedback on value-dependent weighting. These effects may be localized in the accuracy of the subjective scale values or in changes in weight parameters.

One group was given correct feedback using the numerical average (NA) and another group was given no feedback (NF). A comparison of these groups should indicate fundamental differences between feedback and no-feedback conditions. Furthermore, these groups are necessary control conditions for other groups.

In a second comparison, one group was given feedback with respect to a primacy weighted average (PWF) and another group was given feedback with respect to a recency weighted average (RWF). Since *S*s averaged only pairs of distances, a primacy effect is defined as greater weighting on the first stimulus and a recency effect is defined as greater weighting on the second stimulus. The ratio of the order-dependent weights on the feedback trials was set at 2:1 for Group PWF and 1:2 for Group RWF, according to the model in Equation 1. The observed weights may be compared on both feedback and test trials.

In a third comparison, feedback from a value-dependent weighted average was provided where longer distances received greater weight for one group (Group LWF) and where shorter distances received greater weight for another group (Group SWF).

Since the results of the 3 experimental manipulations may follow either an order-dependent weighted average or a value-dependent weighted average model, some procedure is needed to test the specific form of the model. One procedure is to use a general model that includes the other models as special cases. Then the specific models can be statistically compared with the general model. A general model which includes the order-dependent weighted average model, the value-dependent weighted average model, and the simple numerical average model as special cases is

$$R_{ij} = \frac{w v_i s_i + (1 - w) v_j s_j}{w v_i + (1 - w) v_j}, \qquad [3]$$

where w is a weight applied to the first stimulus in the pair, $1 - w$ is a weight applied to the second stimulus in the pair, and v_i is the weight component for the scale value s_i. The sum of all value-dependent weight components is set equal to 1. If the order-dependent weights are equal, Equation 3 reduces to a value-dependent weighted average model; and if the value-dependent weights are equal, it reduces to an order-dependent weighted average model. If both sets of weights are equal, it reduces to a simple numerical average.

Method

Design

The *S*s produced averages of pairs of fixed movements (stimuli, S_1, S_2), where each length was 20, 40, or 60 cm. during pre- and postfeedback trials

TABLE 1

Values of Averages Used for Feedback Groups

S_1	NA			PWF			RWF			SWF			LWF		
	$S_2=15$	$S_2=30$	$S_2=45$	$S_2=15$	$S_2=30$	$S_2=45$	$S_2=15$	$S_2=30$	$S_2=45$	$S_2=15$	$S_2=30$	$S_2=45$	$S_2=15$	$S_2=30$	$S_2=45$
15	15.0	22.5	30.0	15.0	20.0	25.0	15.0	25.0	35.0	15.0	21.0	22.5	15.0	25.0	37.5
30	22.5	30.0	37.5	25.0	30.0	35.0	20.0	30.0	40.0	21.0	30.0	35.0	25.0	30.0	39.0
45	30.0	37.5	45.0	35.0	40.0	45.0	25.0	35.0	45.0	22.5	35.0	45.0	37.5	39.0	45.0

Note. Abbreviations: NA = numerical average, PWF = primacy weighted average, RWF = recency weighted average, SWF = shorter distances receiving greater weight, LWF = longer distances receiving greater weight, S_1 = stimulus of first distance, S_2 = stimulus of a second distance, and 15, 30, and 45 refer to distances in centimeters.

and 15, 30, or 45 cm. during feedback trials. All 9 pair combinations were repeated 3 times in a randomized blocks presentation to obtain prefeedback baseline data. Four randomized blocks of the 9 pair combinations were presented for feedback trials and 3 blocks for postfeedback trials.

During feedback trials, *S*s were told the distance (rounded off to the nearest centimeter) that each response was over or under the average, according to the prescribed model. Feedback given was appropriate to the numerical average for Group NA, a primacy weighted average where $w = 2/3$ for Group PWF, a recency weighted average where $w = 1/3$ for Group RWF, and a value-dependent weighted average (see Equation 2) where $v_{15} = 1/2$, $v_{30} = 1/3$, and $v_{45} = 1/6$ for Group SWF, and where $v_{15} = 1/6$, $v_{30} = 1/3$, and $v_{45} = 1/2$ for Group LWF. The prescribed averages for each condition are listed in Table 1. Group NF produced averages during this stage, but received no feedback.

Apparatus

The apparatus was the same as that used by Levin et al. (1971). Two free-moving, 5-cm.-sq. Plexiglas blocks were mounted on a wooden meter stick. The right-hand block could be locked in place by *E* when defining the length of the fixed movements and could be moved out of the way when *S* was making his averaging response. An additional sliding scale was added for the present experiment so that *E* could read directly the distance that *S*'s response was over or under the average prescribed by the feedback requirements. This scale was zeroed in on the prescribed average before *S* made his averaging response. The apparatus was placed on a table parallel to the coronal plane, and a shield was placed over the meter stick. The *S* was blindfolded so that responding could not be based on visual cues.

Procedure

On each trial, *S* was required to complete the fixed movements S_1 and S_2 and then move the sliding block a distance which he estimated to be the average of the 2 fixed movements. Details of this procedure are given by Levin et al. (1971). The time to complete each fixed movement and return to the starting position was approximately 8 sec. Five seconds were allowed for *S* to produce the average before he was to hold the block steady so that the distance could be read and the feedback given during the feedback trials. Responses were recorded to the nearest millimeter. The distance (rounded off to the nearest centimeter) over or under the prescribed average was then read off by *E*. For example, if the prescribed average was 37.5 cm. and the response was 40.1, *S* heard "over by 3 cm." The interval between trials was 5 sec.

During the feedback trials, *S* was told that he would learn a new type of average that might or might not differ from the "regular" average. This new average was called the "navigational average." The *S* was instructed to use this feedback to influence his subsequent responses so that he could "home in on navigational averages." Two 5-min. breaks were inserted—one following prefeedback trials and one following feedback trials. Finally, a short questionnaire was administered. The *S*s were asked to explain in their own words what the navigational average was, how they made the averages, and if they used any special strategy.

Subjects. Each of the 6 groups was composed of 10 *S*s (5 female and 5 male) recruited from introductory psychology classes at the University of Iowa.

Results

The mean averaging responses on the pre- and postfeedback trials are graphed in the successive panels of Figure 1 for each group.

All main effects of S_1 and S_2 were highly significant for all groups at all stages. The steepness of the lines reflects the effect or weighting of S_2 in determining the average, and the degree of separation of the lines reflects the effect or weighting of S_1. Deviations from parallelism indicate the extent of value-dependent weighting. Lines that converge with increased stimulus magni-

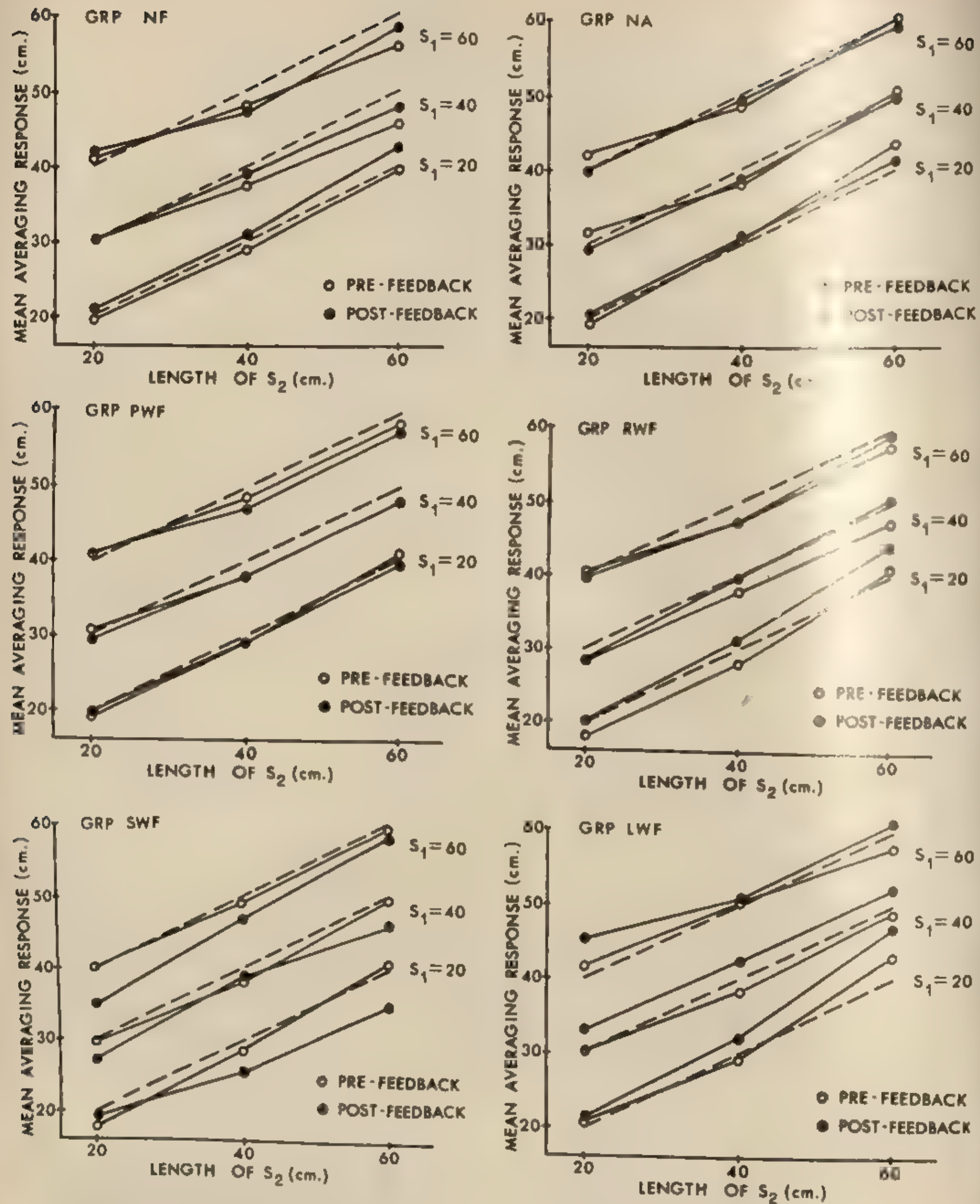


FIGURE 1. Mean averaging responses on pre- and postfeedback trials as a function of the magnitude of the stimuli (S_1 and S_2) for each group. (Dotted lines indicate physical means. Abbreviations: NF = no feedback, NA = numerical average, PWF = primacy weighted average, RWF = recency weighted average, SWF = shorter distance receiving greater weight, and LWF = longer distance receiving greater weight.)

tude imply a greater weighting of longer distances, and lines that diverge with increased stimulus magnitude imply a greater weighting of shorter distances.

Prefeedback Trials

On prefeedback trials, a very slight converging interaction can be noted for most of the groups. The difference between the

pairs 60–20 cm. and 60–60 cm. is less than the difference between 20–20 cm. and 20–60 cm. The form of this interaction is consistent with the interpretation that longer distances received greater weighting. When prefeedback trials were analyzed for initial group differences, all between-groups effects were nonsignificant. The $S_1 \times S_2$ interaction was highly significant, F (4, 216) = 22.52, $p < .001$. Furthermore, this interaction occurred in a large majority of the individual *S* analyses.

Learning Curves on Feedback Trials

Learning curves were constructed for the 4 feedback blocks on Stage 2. Mean absolute error with respect to the average designated as correct was used as an overall measure of learning for the feedback groups. For Group NF, mean absolute error was taken with respect to the numerical average. Figure 2 shows these curves. No systematic changes occurred for Groups NF or NA, through NA tended to be more accurate than NF. No decrease occurred for Groups PWF or RWF; however, mean absolute error for Groups SWF and LWF showed trends to decrease with blocks of trials. With mean absolute error as the dependent variable, the blocks effect was significant for Group LWF, F (3, 27) = 3.72, $p < .05$, but not for Group SWF. It should be noted that it is not at all clear that performance stabilized by the end of the feedback stage. This fact limits the interpretation of data on the postfeedback trials to the levels of learning evidenced on the feedback trials.

Postfeedback Trials

Responses for Group NF did not change from pre- to postfeedback trials. For Group NA, the converging $S_1 \times S_2$ interaction observed on prefeedback trials did not occur on postfeedback trials. This change would be expected due to feedback.

No significant changes occurred on the postfeedback stage for Group PWF. Had *S*s learned and used the primacy weighted average on postfeedback trials, the effect of S_1 should have increased from pre- to postfeedback trials and the effects of S_2 should have decreased. This did not happen. Similarly, for Group RWF, if *S*s used a recency weighted average on postfeedback trials, the effect of S_2 would have increased and the effect of S_1 would have decreased from pre- to postfeedback trials. Though there were trends in these directions, they were not significant. A post hoc analysis for Group RWF compared the mean of the pairs 20–40 cm., 20–60 cm., and 40–60 cm. with the mean of the pairs 40–20 cm., 60–20 cm., and 60–40 cm. The mean of the former pairs was greater than the mean of the latter, indicating recency, t (9) = 2.31, $p < .05$. However, only 2 of the 10 *S*s showed marked effects in the predicted direction.

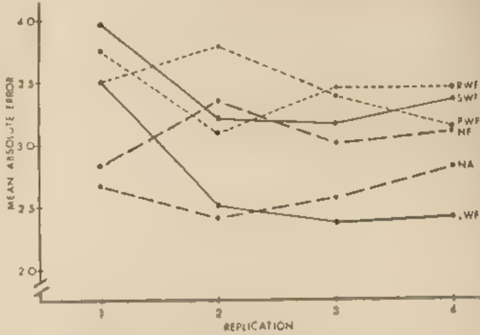


FIGURE 2. Mean absolute error as a function of replications of control trials for the no-feedback (NF) group and replications of feedback trials for the following groups: numerical average (NA), primacy weighted average (PWF), recency weighted average (RWF), shorter distance receiving greater weight (SWF), and longer distance receiving greater weight (LWF).

As seen in the Group SWF panel in Figure 1, a large interaction occurred on postfeedback trials. The form of this interaction differs from the converging interaction noted on prefeedback trials. The lines tend to diverge as stimulus magnitude increases. This interaction is consistent with an integration model that weights shorter distances more than longer distances. The 3-way $S_1 \times S_2 \times$ Stage interaction was highly significant, F (4, 36) = 10.89, $p < .001$. Individual *S* analyses showed that this change in the direction of the interaction occurred for most *S*s (8/10) and reached significance for 3 *S*s ($p < .05$). The overall mean averaging response de-

TABLE 2

Maximum Likelihood Ratio Tests for the General and Restricted Models

	Prefeedback all groups	Group NF	Group PWF	Group RWF	Group SWF	Group LWF
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	11.80	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	(2) 34.50	(2) 17.58	[illegible]

[illegible] square relative to the cell [illegible]. The numbers in parentheses are the degrees of freedom of each [illegible] LWF = longer distances receive [illegible] PWF = primacy weighted average, RWF = recency weighted average, [illegible] weights, and [illegible] order-dependent weight.

creased from 38.8 cm. on prefeedback trials to 36.6 cm. on postfeedback trials, $F(1, 9) = 8.93$, $p < .05$. The value-dependent weighted average model predicts that the averages of heterogeneous pairs should decrease while homogeneous pairs should remain constant from pre- to postfeedback trials. Therefore, a post hoc comparison was made on the homogeneous pairs. No change in homogeneous pairs occurred with stage, $F(1, 9) < 1$.

For Group LWF, the converging interaction increased only slightly from pre- to postfeedback trials, but this trend did not reach significance. The majority of *Ss* showed no change with stage. However, the overall mean increased from 40.0 cm. on prefeedback trials to 42.8 cm. on postfeedback trials, $F(1, 9) = 19.87$, $p < .01$. The value-dependent weighted average model with greater weight for longer distances predicts that responses to heterogeneous pairs should increase while responses to homogeneous pairs should remain the same from pre- to postfeedback trials. Again a post hoc test was made to see if any changes occurred for the homogeneous pairs alone. A significant increase was found of about the same magnitude as the overall increase (2.8 cm.), $F(1, 9) = 35.69$, $p < .001$. Thus, it appeared that feedback did not increase the value-dependent weighting of longer distances as much as it caused an increase in the absolute magnitude of the averaging response.

Model Analysis

The general model (Equation 3) provides for both value-dependent and order-dependent weighting. As noted earlier, these parameters may be systematically restricted to simplify the model. Parameter estimates for these models were obtained by the method of maximum likelihood with the assumption that observations were normally distributed with heterogeneous variances.

Given maximum likelihood estimates, the restricted models may be statistically compared to the general model by likelihood ratio tests (Kendall & Stuart, 1961). The tests for each model are not indepen-

TABLE 3

Parameter Estimates for Best Model

Parameter	Prefeedback trials (all groups)	Postfeedback trials					
		NF	NA	PWF	RWF	SWF	LWF
s_{20}	19.15	20.79	20.89	19.79	20.31	18.91	[illegible]
s_{40}	37.89	38.69	39.05	37.95	39.40	38.05	[illegible]
s_{60}	58.07	57.33	59.91	57.87	59.44	[illegible]	[illegible]
v_{20}	.281	.272				[illegible]	[illegible]
v_{40}	.347	.350				[illegible]	[illegible]
v_{60}	.372	.378				[illegible]	[illegible]
w					.429		
$1-w$					.571		

Note. The parameter $v_{60} = 1 - v_{20} - v_{40}$. Abbreviations: NF = no feedback, NA = numerical a[illegible], PWF = [illegible] weighted average, RWF = [illegible], SWF = [illegible] receiving greater weight, LWF = longer distances receiving greater weight, s = subjective stimulus values, v = value-dependent weight, w = order-dependent weight, and 20, 40, and 60 refer to distances in centimeters.

dent of each other. However, if all possible tests of restrictions are conducted to form a matrix of results, each restriction can be evaluated in combination with all other parameter restrictions. The likelihood function was maximized for the data collapsed over replications and groups on prefeedback trials and for the data collapsed over replications on postfeedback trials for each group.

In all cases, the general model fit the data extremely well. Theoretical means did not deviate appreciably from the observed means. Results of the likelihood ratio tests are presented in Table 2.

Prefeedback trials. The restricted model with $w = 1/2$, implying a lack of order effects, does not fit significantly worse than models including an order-dependent weight. This is indicated in Table 2 by the fact that values in the column for restricted weights are not significant relative to Column 1. However, value-dependent weights are a necessary part of the model as indicated by the large increase in cell entries from Row 1 to Row 2. Table 3 lists the parameter estimates for the most parsimonious model not significantly worse than the general model. The value-dependent weights for prefeedback trials covary with the scale values. This indicates that *Ss* weighted longer distances more than shorter distances. The theoretical means for this model did not deviate appreciably from the observed means.

Postfeedback trials. On postfeedback trials for Group NF, the simplest model was the value-dependent weighted average model since entries in Row 2 of Table 2 are much larger than Row 1. The value-dependent weights listed in Table 3 again covary with scale value. Thus, for Group NF the model did not change from pre- to postfeedback trials and the value-dependent weights remained approximately the same.

Table 2 indicates that for Group NA, both the value-dependent and order-dependent weights are not necessary assumptions for the model to fit the data. The general model does not fit significantly better than a simple numerical average model. Thus, in this group the feedback reduced value-dependent weighting and helped *Ss* to give responses appropriate to the correct numerical average. Though a change in the model from pre- to postfeedback trials occurred, the preceding analysis does not constitute a test of the significance of such a change.

Table 2 indicates that for Group PWF, as for Group NA, the simple average model fits the data no worse than the general model. Feedback did not affect order-dependent weighting as expected; hence, it was not a necessary requirement in the model. For Group RWF, large and significant entries in Table 2 in Column 2 indicate that order-dependent weighting is a necessary requirement of the model. However value-dependent weighting is not necessary. Table 3 gives the parameter estimates for this order-dependent weighted average

model. The value of w indicates that the second (most recent) stimulus received a greater weighting. Thus, the feedback apparently influenced S's responses in this group to weight the second distance more. It should be noted again that the group effect here is primarily localized in the data of a few Ss.

For Groups SWF and LWF, large entries occur in Table 2 when the value-dependent weights are constrained to be equal, but not when the order-dependent weighting is restricted. This indicates that a value-dependent weighted average model is necessary to fit the data for these groups. The value-dependent weights listed in Table 3 indicate that shorter distances received greater weighting than longer distances for Group SWF. This model is in line with the feedback given to this group and is consistent with the highly significant diverging interaction that was obtained on postfeedback trials. Unlike Group SWF, the longer distances received the greater weighting in Table 3, as prescribed by the feedback. These weights are more extreme than those on prefeedback trials, but, as noted earlier, the converging $S_1 \times S_2$ interaction did not increase significantly.

Subjective Scale Values

The subjective scale values estimated for pre- and postfeedback trials for each group are also listed in Table 3. Overall, these values are close to the physical distances although underestimation occurred for 40- and 60-cm. distances. Very little change in the scale values occurred from pre- to postfeedback trials except for Group LWF. As noted earlier, this group tended to increase all responses on postfeedback trials. This is reflected in the increased scale values for Group LWF.

All the stimulus scales, while differing slightly, should be at least linearly related for scale invariance to hold (Anderson, 1970). Likelihood ratio tests were conducted for each group comparing the hypothesis that different scales were needed for pre- and postfeedback trials with the null hypothesis that the scales were linearly related. In all cases the null hypothesis of scale invariance held as indicated by the small and nonsignificant chi-square values.

Discussion

Previous research has shown that the integration function often depends on the particular task. In most experiments, task characteristics have generally remained constant, and the integration rule that Ss use is static. Feedback characteristics had a great effect in the present study on averaging responses, and complex changes in the response pattern resulted. These results indicate that the integration rule may also be a function of feedback and practice, and that the combinatorial rule may be learned. The study of this learning process may help to (a) explain why certain integration rules are used, and (b) delineate the cognitive processes by which Ss actually combine stimuli.

As found in previous studies (Levin et al., 1971; Levin et al., 1973), no order effect obtained with only 2 stimuli on prefeedback trials. Similarly on postfeedback trials for each group except RWF, order-dependent weighting did not obtain. In Group RWF, order-dependent weights were necessary and indicated a greater weighting of the most recent stimulus as predicted by the feedback to this group. Individual S analyses, however, indicated that this effect was evidenced in only a few Ss. On prefeedback trials, value-dependent weights varied directly with stimulus magnitude. This is consistent with the form of the $S_1 \times S_2$ interaction found on these trials. On postfeedback trials, it was found that in Groups NA, PWF, and RWF, value-dependent weights could be eliminated from the model. This was not true for Groups NF, SWF, and LWF. Value-dependent weighting did not change appreciably for Group NF from pre- to postfeedback trials. For Group LWF, the weights which varied directly with stimulus magnitude became a little more extreme, while for Group SWF, the value-dependent weights changed appreciably from directly varying with stimulus magnitude on prefeedback trials to indirectly varying on postfeedback trials.

In the present experiment, one may consider the 2 distances, S_1 and S_2, to be the only stimuli that Ss integrated on each trial. In other tasks to which integration theory has been applied, such as probability learning, feedback over a number of trials constitutes the set of stimuli to be combined (e.g., Freid-

man, Carterette, & Anderson, 1968; Levin, Dulberg, Dooley, & Hinrichs, 1972). In the present task, feedback on previous trials may be viewed as yet another set of stimuli to be integrated along with S_1 and S_2. This process may take on several forms. Each averaging response is a function of S_1, S_2, and feedback on the previous trials. One model may assume that *S* integrates S_1 and S_2 as he did on the first trial and then adjusts or biases his response as a function of previous feedback. Models like this one *bias* the integrated response; they do not change the *rule* by which S_1 and S_2 are integrated.

Other models may assume that feedback alters the integration rule itself or at least the parameters in that rule. For a value-dependent weighted average model like Equation 3, the weights may be a function of feedback. These 2 types of models represent the fundamental psychological distinction between response bias and integration processes. While these 2 processes are conceptually distinct in operation, they are often confounded and inseparable. As far as the integration rule represents a cognitive process, it is encouraging when *S*s report a specific model or process that they used. This, however, does not guarantee that they did indeed use that model. Research on the specific effects of feedback may help to separate the effects of bias from the integration process.

In some cases, feedback appeared to change the form of the integration rule and not response bias when responses to different stimulus pairs changed in different ways. In Groups NA and SWF, the responses to heterogeneous pairs decreased from pre- to postfeedback trials while the responses to homogeneous pairs remained constant. On the other hand, for Group LWF, responses to both heterogeneous and homogeneous pairs increased in magnitude. Feedback affected responses to all the pairs in the same way, and the integration rule did not change substantially as a function of feedback. Feedback either tended to augment the subjective scale values or cause a bias in the response function. Such a result may be due to a possible generalization of feedback to averages of homogeneous stimulus pairs. The *S*s in Group LWF were told that they were under the prescribed average a large percentage of the time. Since initially homogeneous pairs were underestimated, *S*s may have tended to increase all averaging responses. This generalization effect did not occur for Group SWF. In this group, *S*s were almost always told that responses to heterogeneous pairs were over the prescribed average while responses to homogeneous averages were usually under because of the tendency toward underestimation.

Feedback for Groups PWF and RWF had little or no effect on changing the integration rule on the feedback trials. The number of explanations for a null effect are of course unlimited. However, one plausible hypothesis again is that feedback may have generalized between distance pairs. For example, in Group PWF, *S*s were usually told that their responses were over for the 20–60-cm. pair and under for the 60–20-cm. pair. Feedback generalization would functionally lead to the numerical average. Indeed, the pattern of responses for these 2 groups was most like that for Group NA. Several *S*s in Groups PWF and RWF reported that they thought there was a random variation in the new average they were to learn but that it was otherwise the same as the numerical average. Thus, *S*s generally did not discover that order was a relevant cue in the integration function.

Integration rules may be viewed as conceptual rules (e.g., Adams, 1953; Summers & Hammond, 1966; Uhl, 1963). Such conceptual rules are different from those that deal only with the presence or absence of discrete attributes since the components are continuous. The conceptual rule used by *S* may be modeled by a mathematical function which adequately describes the relationship between the stimulus sets and the matrix of responses. While *S*s in Group SWF clearly learned a new conceptual rule, and indicated this verbally following the experiment, *S*s in other groups did not. Similar to studies on rule learning (e.g., Conant & Trabasso, 1964; Hunt & Hovland, 1960), this evidence may indicate that some conceptual rules are learned with greater ease than others. This may be because some aspects of the stimulus set are viewed as naturally contributing to the integration rule while others are not. For example, it is very uncommon for sequential order to contribute to a measure of central tendency.

In conclusion, the present results are evidence that dynamic processes occur in stimulus integration as a function of feedback. Feedback may have a number of effects on the integration rule in terms of weighting parameters or response bias. A careful distinction must be made between the dynamic processes of response bias and integration rule learning similar to the distinction in psychophysics

between decision and sensory processes. The present study supports the notion that integration rules are concepts and as such they may be learned and modified.

REFERENCES

ADAMS, J. Concepts as operators. *Psychological Review*, 1953, **60**, 241–251.

ANDERSON, N. H. Application of an additive model to impression formation. *Science*, 1962, **138** 817 818.

ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170.

ANDERSON, N. H. Looking for configurality in clinical judgment. *Psychological Bulletin*, 1972, **78**, 93–102.

CONANT, M. B., & TRABASSO, T. Conjunctive and disjunctive concept formation under equal-information conditions. *Journal of Experimental Psychology*, 1964, **67**, 250 255.

FREIDMAN, M. P., CARTERETTE, E. C., & ANDERSON, N. H. Long-term probability learning with a random schedule of reinforcement. *Journal of Experimental Psychology*, 1968, **78**, 442–455.

HUNT, E. B., & HOVLAND, C. I. Order of consideration of different types of concepts. *Journal of Experimental Psychology*, 1960, **59**, 220–225.

KENDALL, M. G., & STU[illegible] A. *The advanced theory of statistics.* Vol. 2 [illegible]: Griffin, 1961.

LEVIN, I. P., CRAFT, J. L., [illegible] NORMAN, K. L. [illegible] Averaging of motor movem[illegible]ts of an additive model. *Journal of Ex[illegible]ntal Psychology*, 1971, **91**, 229–235.

Levin, I. P., DULBERG, C. S[illegible], J. F., & HINRICHS, J. V. Sequ[illegible] dependencies in single-item and multiple-i[illegible]bability learning. *Journal of Experime[illegible]hology*, 1972, **93**, 262 267.

LEVIN, I. P., NORMAN, K. [illegible] DOLEZAL, J. M. Response scale effects and [illegible] on processes in the averaging of moto[illegible]ts. *Journal of Motor Behavior*, 1973, 5, [illegible]

SLOVIC, P., & LICHTENSTEIN[illegible] [illegible]mparison of Bayesian and regression [illegible] to the study of information processing [illegible]ent. In L. Rappoport & D. A. Sum[illegible], *Human judgment and social interact[illegible]* [illegible] York: Holt, Rinehart & Winston, 1972.

SUMMERS, D. A., & HAMMON[illegible] Inference behavior in multiple-cue t[illegible]olving both linear and nonlinear relati[illegible]nal of *Experimental Psychology*, 1966, [illegible]57.

UHL, C. N. Learning interval[illegible]: I. Effects of differences in stimulus w[illegible] *Journal of Experimental Psychology*, 196[illegible] 273.

(Received March 5[illegible]

Journal of Experimental Psychology
4, Vol. 102, No. 3, 409–414

VISUAL IMAGERY FOR WORDS: THE HEBB TEST[1]

ROBERT J. WEBER[2] AND ROGER HARNISH

Oklahoma State University

Two experiments were conducted to examine Hebb's test of the "picture theory" of visually imagining words. Experiment I examined retrieval from visual image representations of words by using a method of probing for the spatial properties of the *i*th letter in an imagined word. Evidence was obtained for a limited-capacity visual image system that can represent 3- and 5-letter words, in a spatially parallel manner, as efficiently as they can be represented in a visual percept system. Experiment II examined the generation of visual image representations of words. All parts of a word were not generated simultaneously, and generation generally did not proceed from left to right. The results of the 2 experiments were interpreted as indicating the existence of a visual image operating memory of limited letter capacity. The contents of the memory can be examined in much the same way as visual percepts.

Hebb (1966, 1968, 1972) and Woodworth and Schlosberg (1954) describe at length what we might term the "picture theory" of visual imagery. According to this theory, a self-generated mental picture or image can be examined in much the same way as a physical picture is examined by perception. For example, many people report visual images of whole lines or stanzas of verse, or at least of single words. The presumption is that these images are spatially parallel, i.e., all words or letters in the composite image are simultaneously available. Hebb proposes a test of the picture theory. The test contains a general requirement that there exist analogs between perception and imagery for the processing of spatial information, and a particular requirement that these analogs reveal themselves in a spelling test. To experimentally verify the picture theory, Hebb feels it is necessary to show that *S*s can spell imagined words in a backward direction as fast as they can in a forward direction. Since he finds that the backward spelling of imagined words is much slower than the forward spelling, Hebb (1966) concludes that whatever the nature of a visual memory image, it is not like having a mental picture which can be examined by the "mind's eye":

> the subjective impression that one can "look at" one's image freely is shown by objective test to be wrong [pp. 43–44]. The fact that the person with visual imagery can only "see" the letters of a word in left-to-right order shows clearly that the memory image . . . is a *series* of events in a *particular order*, not a picture whose parts could be looked at in any order (or simultaneously) [p. 46; italics added].

It is the contention of this article that, with qualifications, the picture theory of visual imagery is at least partly correct, and that Hebb's (1966) general requirement (that there exist analogs between percept and image processing) is a perfectly reasonable test of the theory. However, we feel that Hebb's particular requirement, a comparison of forward and backward spelling times, is not appropriate, for the following reasons. First, a spelling test is not necessarily visual. The *S* may simply draw on a highly practiced and extensively trained verbal/speech representation of the word when spelling it. Second, the spelling test may involve an assymetrical use of verbal and visual image representations. The *S* may indeed have a visual image representa-

[1] The research reported herein was supported in part by a grant from the Research Foundation, Oklahoma State University, and in part by a grant from the U.S. Office of Education, Department of Health, Education, and Welfare, both to the first author. The opinions expressed do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement should be inferred.

[2] Requests for reprints should be sent to Robert J. Weber, Department of Psychology, Oklahoma State University, Stillwater, Oklahoma 74074.

tion of the word but may still use his familiar verbal/speech representation for spelling in a forward direction, while possibly dropping back to his visual representation for the unfamiliar backward spelling. The more rapid forward spelling would then simply follow from the fact that, for sequential items, speech is a more rapid process than visual imagery (Weber & Castleman, 1970). Third, the visual perception system may have a substantially greater spatially parallel processing capacity than the visual imagery system. A suggestion that the visual imagery system may be of very limited capacity has been put forth by Weber, Kelley, and Little (1972). If visual imagery is of more limited capacity than visual perception, a spelling test using long words, e.g., *university*, may be biased in favor of perceptual performance. Fourth, even if capacity limitations were equal for the image and percept systems, the way in which items (letters) sequenced in and out of the 2 systems might well differ (Weber et al., 1972).

In the present study, an attempt was made to eliminate some of the difficulties associated with Hebb's (1966) spelling test. First, to make more certain that *S*, when requested to visually imagine a word, does indeed use a visual representation, these experiments required *S* to respond to visual spatial properties of the letters in a word. In particular, we asked *S* to imagine words with lowercase printed letters and to classify a given letter for the spatial property of vertical height. Those letters that are vertically large (b, d, f, g, . . ., y) would fall in the *yes* category and all other letters would fall in the *no* category (Weber & Castleman, 1970; Weber & Kelley, 1972; Weber et al., 1972). This procedure allows for a set of converging factors which point to the conclusion that a visual image system rather than a verbal system is involved. The converging factors are: (*a*) the ability of *S*s to correctly classify the imagined letters for vertical height, (*b*) frequent subjective reports from *S*s that they do visualize during the task, and (*c*) the finding of highly distinctive rates of processing for visual image [illegible] speech instructional sets ([illegible] Castleman, 1970). Second, to [illegible] maximum parallel representation [illegible] ual image and to avoid building [illegible] essary sequential processing, a [illegible] ort probe technique was used [illegible] ent studies. In this technique, [illegible] osition in a word is indicated by [illegible] robe, with probe position define[illegible] to right. The letter correspo[illegible] probed position is then classifi[illegible] d height by *S*. The probe te[illegible] allows for a direct compari[illegible] ept and image representation [illegible] xamination of capacity effect [illegible] nulus words of different length [illegible] nted.

Experime[illegible] rd Probe

This experi[illegible] oncerned with comparing retri[illegible] om an available image vs. an ava[illegible] ept representation. The follo[illegible] ons were asked: Does response [illegible] for image and percept conditions vary as a function of the number of letters in a word? Do the functions relating RT to probe position differentiate between image and percept representation?

Method

Subjects. Sixteen experimentally naive volunteer *S*s with normal or corrected vision were tested individually. Each *S* was previously screened in a manner similar to that of Weber and Castleman (1970) for ability to respond to a visualized alphabet.

Stimuli and procedure. The test stimuli were nouns and adjectives with a Thorndike–Lorge (1944) count of 5 or more per 100,000 occurrences.[3] A total of 42 different words were used. Twelve were test words of 5 letters in length and 12 were test words of 3 letters in length. In addition, there were 8 filler words 3 letters in length (for statistical balancing purposes) and, finally 5 practice words for each word length. The 12 5-letter test words were selected such that vertically large lowercase letters (b, d, f, g, h, j, . . ., y) and vertically small letters (a, c, e, . . ., z) appeared equally in all 5 letter positions (first through fifth). The 12 3-letter test words and 8 3-letter filler words were also

[3] Strictly speaking, this was true only of the word stem. Occasionally, for example, a 4-letter word might have been used with a plural form (*s* added) in order to satisfy conditions and yield a satisfactory 5-letter word.

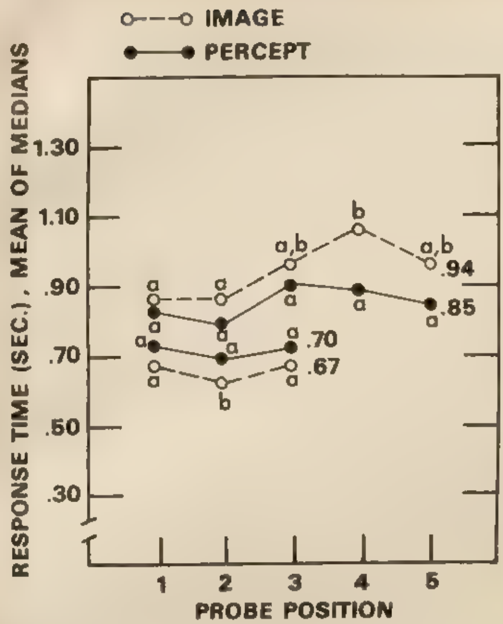


FIGURE 1. Response time as a function of image–percept representation and letter position probed in Experiment I.

selected so that vertically large and small letters appeared equally in all 3 letter positions.

Each S was given a preliminary spelling test for the words that were to be used. He was required to "spell" the words both with letters and with *yes* and *no* (indicating the vertical height of the letters). For example, the word *cat* would be spelled *c*, *a*, *t* and *no*, *no*, *yes*, respectively.

The stimulus words were presented as images or as percepts. For the image conditions, E spoke the word and S was instructed to form a visual image of it with his eyes closed. Imagining something visually was explained to be like "picturing it in the mind." For the percept condition, each word was rear projected with a Kodak Carousel projector equipped with a solenoid-operated shutter. Letters appeared in lowercase elite type on a 30.48 × 20.32 cm. rear-projection screen. The vertically small letters were .63 cm. high. Since vertically large letters were slightly irregular in height, the *j*, 1.27 cm. in projected height, was taken as the standard. The S viewed the letters from a distance of 53.34 cm.

Four seconds after presentation of a stimulus word, E orally presented a probe digit (1–5). The probe digit indicated a letter position in the word S was imagining or perceiving. In a word like *cat*, the digit *1* referred to the *c*, *2* referred to the *a*, and *3* referred to the *t*. The 5-letter words were probed in the same fashion. The spoken digit activated a Lafayette voice relay and started a Standard clock. The S then responded *yes* or *no* to indicate whether the letter denoted by the probe digit was vertically large or not. The S's response stopped the clock, and RT from onset of the probe to onset of the response was recorded. For the percept conditions, after S responded, the slide terminated and a blank adaptation field was displayed from another projector with a solenoid shutter. All Ss were instructed to respond as rapidly as possible, with 100% accuracy. Any words which resulted in error were repeated at the end of the run. Hence, all time scores were for correct choices only.

Design. There were 2 levels of word length (3-letter and 5-letter words) and 2 modes of representation (percept and image). Word length was a between-Ss variable and representation mode was within Ss, with half of the Ss proceeding in the order image–percept and the other half in the order percept–image. There were 60 trials within each of the between-Ss conditions (each test word appeared 5 times in the 5-letter treatment and 3 times in the 3-letter treatment). This allowed each letter position in each word to be tested once and only once; the additional 3-letter filler words made it possible to equate practice on 3- and 5-letter words. Probe position was also treated as a separate within-Ss variable for each word length. Two different random orders of words were used.

Results and Discussion

Each S's median RT for a given condition was computed, and the mean of the medians was then obtained for Figure 1. Thus, each point represents 96 events (12 Words × 8 Ss). Medians were used because of the substantial variability encountered. The number to the right of each function is the mean RT obtained by collapsing across probe positions.

Several statistical analyses were conducted, and RT was the dependent variable in each case. In the first analysis of variance, data were collapsed across probe positions. That is, each S's score was the mean of his separate median probe position RTs. The resulting 2-way classification of 2 word lengths (3 and 5 letters) by 2 representation modes (image and percept) was analyzed, with the first factor between Ss and the second within Ss: for word length, F (1, 14) = 5.20, $p < .05$; for representation mode, F (1, 14) = .48, $p > .05$; and for the Word Length × Representation interaction, F (1, 14) = 2.82, $p > .05$. Thus, RT significantly increases with word length, but images and percepts are not reliably different in the speed with which spatial information is extracted from them. Indeed, processing time from the

3-letter image condition is at least as fast as that from the 3-letter percept condition.

The second series of analyses examined probe position effects. Consider the 3-letter words. Here, the analysis was based on a 2-way classification of 3 probe positions (1, 2, and 3) by 2 representations (image and percept), with both factors within *S*s: for probe position, $F\ (2, 14) = 7.64$, $p < .05$; for representation mode, $F\ (1, 7) = .94$, $p > .05$; and for the Probe Position × Representation interaction, $F\ (2, 14) = .09$, $p > .05$. Thus, the dip in RT for the middle letter is significant, suggesting that all parts of a word are not processed equally rapidly.

To examine this serial position effect more closely, separate Newman–Keuls tests were performed for the 3-letter percept and 3-letter image conditions. The results are indicated by the letters adjacent to each plotted position in Figure 1. Within the range of each function, the probe positions having the same letter do not differ significantly; however, the probe positions having different letters *do* differ significantly ($p < .05$). Hence, the 3-letter percept condition may be regarded as statistically flat, a result consistent with a random access process. But the 3-letter image condition indicates a significantly faster processing time when the second-letter position is probed. However, it is appropriate to reemphasize that the Probe Position × Representation interaction was not significant in the analysis of variance for probe position effects.

In a comparable series of analyses for the 5-letter word conditions, the following results were obtained: for probe position, $F\ (4, 28) = 4.97$, $p < .05$; for representation mode, $F\ (1, 7) = 1.88$, $p > .05$; and for the Position × Representation interaction, $F\ (4, 28) = 2.28$, $.05 < p < .10$. Again, the probe position effect shows that not all parts of a word are processed at the same rate.

The Newman–Keuls test results are indicated in Figure 1 in the manner described previously. The 5-letter percept function is statistically flat, while there appears to be a serial position effect developing for the 5-letter image fun[illegible] However, it should be noted that [illegible] Probe Position × Representation M[illegible] interaction fell short of significanc[illegible] the analysis of variance for 5-letter [illegible] position effects. It is our impression [illegible]he data indicate that *S*s' capacities [illegible]isually imagine letters were strained [illegible] 5-letter image condition. This may [illegible] either that 5 letters cannot be eff[illegible] visualized all at once or that the st[illegible] [illegible] the image is decreased and all or [illegible] fades in and out. The *S*s may co[illegible] for this lost capacity in several w[illegible]ey may verbally mediate some [illegible] the letter images along the line[illegible]ed by Weber et al. (1972). They [illegible] some of the letters in a visual ima[illegible] not all of the letters of a word may [illegible]erated for use at the same time. The data are not sufficient to decide.

It is appropriate to note that preliminary analyses were also performed in which *yes/no* response categories were treated as a separate factor. No consistent effects distinguishable from the present results were apparent.

To summarize, there are differences in processing time between 3- and 5-letter word conditions. There do not seem to be significant differences in processing time for overall image vs. percept representation, but there may be differences in probe position effects. In regard to probe position effects, percept condition processing is consistent with random access to serial positions, but image condition processing may not be indicative of strict random access to serial positions.

Experiment II: Probe Word

Experiment I examined differences in the processing of information from *available* visual images vs. *available* visual percepts. The present experiment is concerned with examining differences in the processing of information from visual images vs. visual percepts, during the generation process. By presenting the probe before the word that is to be imagined or perceived, the necessity for generating an image or percept is added to the task. The temporal growth

of the image or percept can then be examined.

Method

Sixteen experimentally naive *S*s with normal vision were paid for their participation and were tested individually. All stimulus materials and procedures were identical to those of Experiment I except that the probe digit was presented first and then was followed by the stimulus word.

Results and Discussion

The treatment of the results, summarized in Figure 2, was identical to that of Experiment I. When the probe position variable was collapsed, the resulting 2-way analysis of word length by representation mode indicated significant effects of word length, $F\ (1,\ 14) = 7.69$, $p < .05$; representation mode, $F\ (1,\ 14) = 249.58$, $p < .05$; and the Word Length × Representation interaction, $F\ (1,\ 14) = 9.72$, $p < .05$. In Experiment I, only the word length variable was significant. The present results indicate that it takes significantly longer to generate and process an image than it does a percept. The significant Word Length × Representation interaction is indicative of some form of limited capacity in the visual image generating and processing system. As the word to be generated becomes longer and exceeds processing capacity, RT becomes disproportionally longer for the image conditions in comparison to the percept conditions.

Separate serial position analyses were performed for each word length. For 3-letter words, an analysis of probe position by representation mode yielded significant effects of position, $F\ (2,\ 14) = 14.56$, $p < .05$; representation, $F\ (1,\ 7) = 336.78$, $p < .05$; and no significant interaction, $F\ (2,\ 14) = 2.93$, $p > .05$. The Newman–Keuls test for 3-letter word conditions indicated that the percept function is statistically flat and that for the image function each probe position is statistically different from the other 2. Thus, generating percepts appears to be an all-at-once parallel process, and the generation appears to have no effect on the processing of information from the percept. The increasing 3-letter image function is consistent with a left-to-right generating and/or processing strategy.

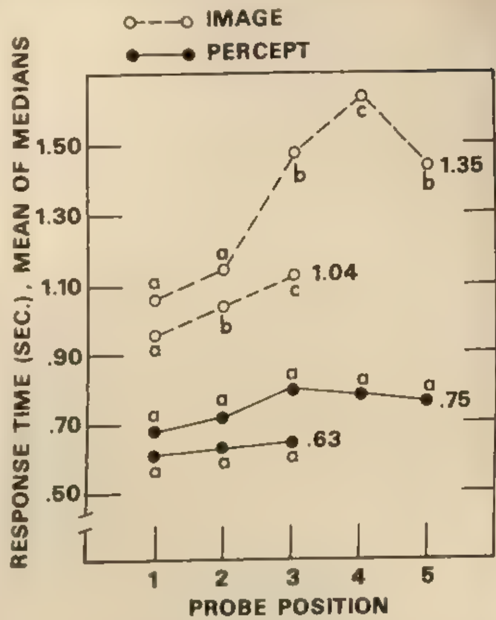


FIGURE 2. Response time as a function of image–percept representation and letter position probed in Experiment II.

The serial position analyses for 5-letter words indicated the following. In the probe position by representation analysis, the effects of position, $F\ (4,\ 28) = 26.76$, $p < .05$; representation $F\ (1,\ 7) = 99.06$, $p < .05$; and the Position × Representation interaction, $F\ (4, 28) = 13.77$, $p < .05$, were significant. The Newman–Keuls test indicated that the percept function is statistically flat and that the image function has an inverted-U form. These results are not consistent with a left-to-right generating and/or processing strategy.

In summary, percept functions remain statistically flat even when the probe is presented prior to the word. Generation processes of the type studied here seem minimal in perception, and there still appears to be random access to the different serial positions. The random-access process could occur in at least 2 different ways. First, the percept representation could be of a spatially parallel nature in which all the letters are simultaneously available for processing. Second, after receiving the probe digit, *S* could simply visually fixate

on the appropriate letter location on the viewing screen in anticipation of the forthcoming word. The latter process would not be very interesting theoretically. Moreover, it would not explain the flat percept functions of Experiment I, where the locus of the probe position was uncertain.

General Discussion

The picture theory of visual imagery was supported in Experiment I; Hebb's (1966) general requirement that performance analogues exist for percept and image representation was satisfied. Access to a 3-letter image representation was at least as rapid as to a 3-letter percept representation. Access to a 5-letter image representation was not significantly slower than to a 5-letter percept representation. However, there was some suggestion of an emerging Percept–Image Representation × Word Length interaction. This interaction was strongly present in Experiment II, which dealt with image generation time.

Our explanation for these findings is that there exists a visual image operating memory with a fixed letter capacity for parallel processing that is less than that of the visual percept system. When the image capacity of the operating memory is strained or exceeded, differences in processing time between percept and image systems become apparent. These differences reveal themselves in effects of word length and/or serial position. The effects were more apparent in Experiment II, where the image had to be both generated and scanned, than in Experiment I, where the image could have been generated and available some time before scanning began. As a partial independent check on this capacity interpretation, we have since performed several scaling studies of *S*s' subjective impressions of how many letters they can visualize at one time. The median number of alphabetic letters simultaneously visualizable was found to be 5, with an ascending method of limits procedure ($n = 31$), and 6, with a numerical estimate procedure ($n = 24$). These scaling values seem very supportive of the results from Experiments I and II.

If the visual image system does have a capacity limitation, it is of interest to know how this limitation is handled. Weber and Kelley (1972) and Weber et al. (1972) have found evidence that when an array of letters such as an alphabetic list or long word (e.g., "supercalifragilisticexpialidoccious") exceeds the spatially parallel capacity of the visual image system, a verbal sequencing comes into play. Briefly, what happens is that *S* says to himself each successive letter or syllable name prior to visually imagining it. This verbal sequencing allows for the sequential ordering and generation of visual images and is in sharp contrast to perceptual presentation, where *S* need only visually scan or move his eyes from one letter or syllable locus to another (Weber & Kelley, 1972) in order to sequence through a list. Thus, the capacity of the visual percept system is essentially unlimited as long as the time requirements are not so stringent as to limit visual scanning and/or eye movements. In this view, the mode of sequencing between letters or syllables seems to be one fundamental distinction between the visual image and visual percept systems.

In conclusion, the majority of the findings of the present study indicate that the picture theory of visual imagery may have much more to recommend it than has previously been supposed (Pylyshyn, 1973). This is not to say, however, that there are no differences between percept and image systems. One important difference seems to be in the manner of sequencing between items (letters) in a list that exceeds capacity. There are, no doubt, other differences to be discovered.

References

Hebb, D. O. *A textbook of psychology.* (2nd ed.) Philadelphia: Saunders, 1966.

Hebb, D. O. Concerning imagery. *Psychological Review*, 1968, **75**, 466–477.

Hebb, D. O. *A textbook of psychology.* (3rd ed.) Philadelphia: Saunders, 1972.

Pylyshyn, A. What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 1973, **80**, 1–24.

Thorndike, E. L., & Lorge, I. *The teacher's word-book of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

Weber, R. J., & Castleman, J. The time it takes to imagine. *Perception & Psychophysics*, 1970, **8**, 165–168.

Weber, R. J., & Kelley, J. Aspects of visual and acoustic imagery. *Psychonomic Science*, 1972, **27**, 121–122.

Weber, R. J., Kelley, J., & Little, S. Is visual image sequencing under verbal control? *Journal of Experimental Psychology*, 1972, **96**, 345–362.

Woodworth, R. S., & Schlosberg, H. *Experimental Psychology.* New York: Holt, 1954.

(Received April 13, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 415–422

VERBAL DISCRIMINATION AS A CONCEPT-ATTAINMENT TASK USING THE EVALUATIVE DIMENSION [1]

MARIAN SCHWARTZ [2]

University of Wisconsin—Madison

In Experiment I, a verbal discrimination task was used in which correctness was determined by a relational rule based on the evaluative dimension. Each pair consisted of a neutral word paired with an extreme ("good" or "bad") word. When the word closer to the positive pole of the dimension was correct, performance was significantly better than when correctness was randomly determined. When the word closer to the negative pole was correct, performance was not significantly different from the random control. These results, coupled with the results of a postlearning questionnaire, suggest that most *S*s failed to conceptualize the evaluative dimension as a single bipolar continuum. In Experiment II, it proved harder to judge which member of a pair was closer to the negative end than to judge which was closer to the positive end of the dimension, indicating an asymmetry in rule utilization as well as in rule discovery.

Several recent experiments have found that the evaluative attributes of words can be a basis for concept attainment (Di Vesta & Ingersoll, 1969; Haygood, 1966; Rhine, 1965; Rhine, Cole & Ogilvie, 1968; Taylor & Haygood, 1968) or for differential encoding in memory, as indexed by release from proactive inhibition (Wickens & Clark, 1968). These studies used stimulus words drawn from the positive (good) and negative (bad) halves of the evaluative dimension of connotative meaning. The results have been interpreted (e.g., Haygood, 1966; Wickens & Clark, 1968) as evidence that positive and negative words constitute psychologically distinct classes, and as providing at least partial validation for the evaluative scale of the semantic differential.

According to the bipolar model which underlies the semantic differential (e.g., Osgood, Suci, & Tannenbaum, 1957), each scale is assumed to represent a single bipolar dimension extending as a straight line in semantic space, with the midpoint of the line passing through the origin of this space. As applied to the evaluative dimension, such a model implies not only that positive and negative words should constitute different psychological classes, but that these two classes of words should be perceived (responded to) as belonging to the same dimension; and further, that evaluatively neutral words should be perceived as falling along this same dimension midway between the two extremes.

The present study tested some of the implications of the bipolar model, as it applies to the evaluative dimension, by investigating the attainment of concepts based on relational rules. In Experiment I, a verbal discrimination (VD) paradigm was used to present pairs of words in which one member was evaluatively neutral and the other was either positive or negative. Correctness was determined by one of two rules. Under the positive rule, the correct member of each pair was the one closer to the positive pole of the dimension; under the negative rule, the correct member was the one closer to the negative pole. It was expected that *S* should discover and utilize the appropriate relational rule if, in accordance with the bipolar model, *S* conceptualized positive, neutral, and negative words as being monotonically ordered along the same dimension. More specifically, if the bipolar model is applicable to performance in the

[1] This research was supported in part by a National Institutes of Health Biomedical Sciences Support Grant to the Graduate School of the University of Wisconsin. Experiment I was reported at the Psychonomic Society meeting, St. Louis, November 1971.

[2] Requests for reprints should be sent to Marian Schwartz, Department of Psychology, University of Wisconsin—Madison, Psychology Building, Charter at Johnson, Madison, Wisconsin 53706.

present task, the following outcomes would be expected: (*a*) Where correctness is determined by a relational rule, VD performance should be significantly above the level of a control group in which correctness is randomly determined. (*b*) The positive and negative rules should be equally effective. (*c*) Assuming that *S*s are able to verbalize the rules when questioned at the end of learning, they should tend to express the rules in relational rather than absolute terms. For example, the positive rule could be expressed either as a relative statement such as, "The more positive (closer to good) member of each pair was correct" or in absolute terms, such as, "Where a neutral and a good word were paired, the good word was correct; where a neutral and a bad word were paired, the neutral word was correct." The relational form, but not the absolute form, reflects the idea of a single bipolar dimension ranging from good through neutral to bad.

A secondary question concerned the effect of instructions upon discovery of the appropriate rule. Some *S*s were given an instructional clue to the effect that correctness was rule-determined and that the rule involved evaluative attributes. The presence or absence of this instructional clue was orthogonal to type of rule.

Experiment I

Method

Design. There were five independent groups: positive, instructed-positive, negative, instructed-negative, and random. The positive rule applied to the first two groups; the negative rule, to the next two groups. Both instructed groups received the same instructional clue, and this clue constituted the only difference in treatment between the two instructed groups and their uninstructed counterparts. In the random (control) group, correctness was randomly determined. The random group, therefore, could learn the list only by rote; the four rule-determined groups could learn either by rote, or by making use of an appropriate rule, or by some combination of rule and rote.

Lists. All groups learned a list of 16 pairs in which 8 pairs were neutral-good (i.e., a word rated neutral paired with a word rated good) and 8 were neutral-bad. The construction of the lists might be called semi-double-function. That is, each neutral word occurred in two pairs in the same list (e.g., the pairs AMOUNT, HOPE, and AMOUNT, CRIME), being correct in one pair and incorrect in the other. Subjects cannot master such a list by learning which individual words are correct; they must take the other member of the pair into account. The list was constructed in this way to increase the difficulty of rote learning and to encourage attention to both members of each pair, thereby maximizing *S*'s opportunities to discover the appropriate relational rule.

The occurrence of each neutral word in two pairs while each extreme (good or bad) word occurred in only one pair should introduce a bias in favor of the neutral words, according to the frequency theory of VD learning (Ekstrand, Wallace, & Underwood, 1966). That is, all other things being equal, the accumulation of more frequency units for the neutral words should produce more correct responses for pairs in which neutral words are correct than for pairs in which neutral words are incorrect. A bias favoring neutral words, however, should not differentially affect performance as a function of group, as in all five groups the neutral member was the correct one for half of the pairs.

In selecting the words to be used, it was desired to equate the good, bad, and neutral sets on irrelevant variables, but this was not feasible using published lists of words rated on the evaluative dimension. Accordingly, a set of 224 nouns which appeared by inspection to be unequivocal in their goodness, badness, or neutrality were drawn from the Paivio, Yuille, and Madigan (1968) norms. As a check on *E*'s judgments, these nouns were then rated on a 5-point evaluation scale by 16 students recruited from a course in introductory psychology. Three sets of nouns—8 good, 8 bad, and 8 neutral—were then chosen to meet the following criteria: (*a*) the three sets were equated as far as possible in word length, in number of first-letter duplications, and in the variables of frequency, imagery, and meaningfulness as measured by the scales provided in the Paivio et al. norms, and (*b*) the designation of words as good, bad, or neutral on the basis of *S* ratings should be as unequivocal as possible. The 128 ratings made by the 16 raters for the 8 words in each set break down as follows: for the good set, 97% of the ratings were either Very Good or Somewhat Good, and 3% were Neutral; for the bad set, 97% were either Very Bad or Somewhat Bad, and 3% were Neutral; for the neutral set, 88% of the ratings were Neutral, 6% Somewhat Good, and 6% Somewhat Bad. It is recognized that, because of the rough scaling procedure and small number of scalers, the sample of neutral words selected may not adequately represent the psychological midpoint between the samples of good and bad words selected. If an imbalance of this kind exists, however, it should not bias the comparison between groups, since the same pairs were used for all groups. (This point is covered more fully in the discussion of the results.) The words selected for the good set were FREEDOM, GIFT, HOPE, HUMOR, KINDNESS, MOTHER, TRUCE, VIGOR; for the bad set, ASSAULT, CRIME, DEATH, DECEIT, PYTHON, SLAVE, TROUBLE, VICTIM; and for the neutral set, AMOUNT, ANKLE, FORK, HAIRPIN, INSTANCE, LENGTH, PRODUCT, UNIT.

Eight 16-pair lists were constructed. In each list, the 8 neutral words each occurred in two pairs and the remaining 16 words each occurred in one pair. The eight lists comprised two basic pairings with four versions of each pairing. The four versions differed only with respect to the designations of the correct pair members. In one version, the positive rule applied; i.e., the correct words were the positive members of neutral–good pairs and the neutral members of neutral–bad pairs. In another version, the negative rule applied. Whichever member was correct in the positive version was necessarily incorrect in the negative. There were two random versions of each pairing, which differed in that whichever word was correct in one version was incorrect in the other. In each random version, the neutral members were designated as correct for half the neutral–good pairs and for half the neutral–bad pairs. Within each list, the number of pairs with the shorter word correct equaled the number with the longer word correct.

Four orders of presentation were used for each list. The orders were random subject to certain restrictions, such as that no more than three consecutive pairs could be of the same type (i.e., neutral–good or neutral–bad). The assignment of the correct member to the upper or lower position within a pair was random subject to certain restrictions designed to prevent confounding between intrapair position and any other variable.

A practice VD list of eight pairs of numbers was constructed by pairing the numbers 1–16, with each number occurring in one pair. In the practice list, pairing and correctness were random subject to certain restrictions designed to minimize the possibility that any rule could be used as a cue to correctness. A single pairing, with four orders of presentation, was used for the practice list.

Procedure. Subjects were run individually. The practice list of numbers was presented until *S* reached the criterion of two perfect trials, not necessarily consecutive. The word list was then presented for 12 trials. All lists were shown by the anticipation method on a Stowe memory drum at a 2:2 rate, with an 8-sec. intertrial interval. The same starting order was used for all *S*s. Since the standard VD instructions to call out the correct member of each pair might induce a set toward evaluatively positive words, the instructions to *S* avoided the terms *correct* and *incorrect* (or *right* and *wrong*) in referring to the two members. Instead, instructions for both the practice list and the word list indicated that pairs would appear on the left-hand side of the memory-drum window, and that the task was to anticipate which member of each pair would appear on the right-hand side. The word-list instructions also informed *S* that some words would occur in more than one pair, and that a word which appeared on the right-hand side when it was a member of one pair would not necessarily appear on the right when it was a member of another pair. The only reference to rules or to the evaluative dimension occurred in the following sentence, which was added to the word-list instructions for the instructed-positive and instructed–negative groups: "There is a rule you can discover which will help you learn which words will appear on the right. I cannot tell you the exact rule, but it has to do with the extent to which words arouse reactions of goodness or badness." (The order of the words "goodness" and "badness" in this sentence was reversed for half the *S*s within each of these two groups.)

At the conclusion of the word list, an untimed written questionnaire was administered. The questionnaire instructions informed *S* that different people received different word lists, and asked him to check whichever one of the following three statements seemed to apply to his list: 1. It seemed to be arbitrarily determined which words would appear on the right. 2. There seemed to be a rule or principle determining which words would appear on the right. 3. For some word pairs, it seemed to be arbitrary; for other word pairs, there seemed to be a rule or principle. If *S* checked Statement 2 or 3, he was asked to write the rule.

Subjects. The *S*s were 40 paid summer session students, 20 men and 20 women, none of whom had participated in the original rating of the words to be used. Eight *S*s, 4 men and 4 women, were assigned to each of the five conditions, using a block randomization procedure. The two pairings of the word list were balanced over condition and sex of *S*, and the two versions of each random list were balanced over sex of *S*. Three additional *S*s were eliminated; 2 failed to reach criterion on the practice list within 18 trials, and the other had four failures to respond on the word list.

Results and Discussion

Practice list. Because all *S*s were treated alike on the practice list, performance on that list should not differ among the five groups. Mean trials to criterion ranged from 6.62 to 10.75 for the five groups, and analysis of variance confirmed that these practice-list differences were not significant ($p > .10$). Nonsignificant effects ($p > .10$) were also found for sex of *S* and for the Group × Sex of *S* interaction.

Word lists. Mean correct responses for the 12 word-list trials combined are shown in Table 1, and trial-by-trial data are shown in Figure 1, Panel A. An overall 5 × 2 × 2 analysis of variance was performed on the total number of correct responses. The independent variables were rule (positive, instructed–positive, negative, instructed–negative, and random), sex of *S*, and type of pair (neutral–good vs. neutral–bad). The variable of rule proved to be highly significant, F (4, 30) = 10.14, $p < .001$. Type of pair was also significant, F (1, 30)

TABLE 1

Mean Correct Responses on Word List, by Condition and Type of Pair

Condition	Type of pair: All pairs	Neutral–good	Neutral–bad
Random	148.50	73.50	75.00
Neutral word correct	76.62	37.88	38.75
Extreme word correct	71.88	35.62	36.25
Negative	156.62	81.25	75.38
Instructed–Negative	164.75	84.88	79.88
Positive	177.12	88.75	88.38
Instructed–Positive	189.75	95.25	94.50
All conditions combined	167.35	84.72	82.62

$= 4.95$, $p < .05$, with performance on neutral–good pairs somewhat higher than on neutral–bad. The only other factor approaching significance was the Rule × Type of Pair interaction, $F\ (4,\ 30) = 2.27$, $.05 < p < .10$.

As shown in Table 1 and in Figure 1, Panel A, the five rule groups in order from lowest to highest performance on the word list were random, negative, instructed–negative, positive, and instructed–positive. Dunnett's test was used to compare each of the four rule-determined groups with the random control. With $k = 5$ and $df = 35$, performance was significantly above the random-group level in the positive group, $t = 3.84$, $p < .01$, and in the instructed–positive group, $t = 5.54$, $p < .01$, but not in the negative, $t = 1.09$, $p > .05$, or instructed–negative, $t = 2.18$, $p > .05$.

An additional analysis comparing positive vs. instructed–positive groups revealed a highly significant effect of instructions, $F\ (1,\ 14) = 22.75$, $p < .001$. The analogous negative vs. instructed–negative comparison did not approach significance ($F < 1$).

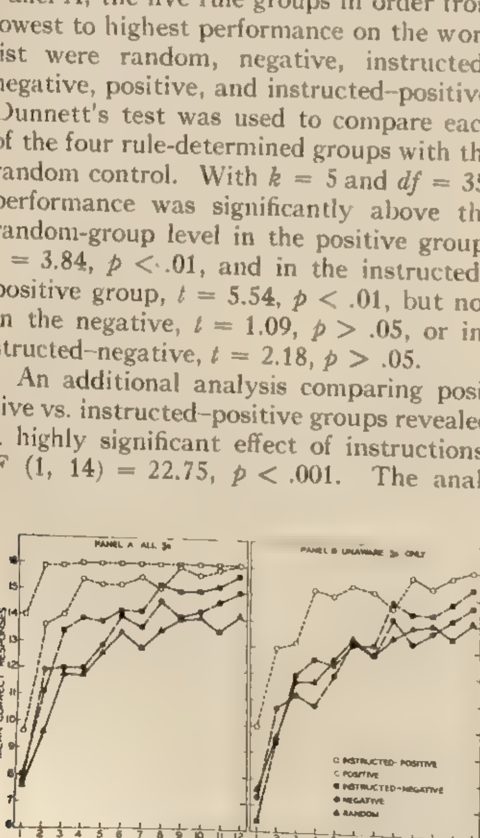


Figure 1. Mean correct responses on word list, by condition.

The major results for the rule variable may be summarized as follows: Where the positive rule governed correctness, performance was significantly better than in the random group, where learning could occur only by rote; where the negative rule governed correctness, performance was slightly but not significantly above the random-group level. These results held both for the instructed and for the uninstructed groups.

The question might be raised whether the superiority of the positive to the negative rule could be accounted for in terms of a simple response bias; i.e., a general tendency to choose positive and or avoid negative words (cf. Boucher & Osgood, 1969). Against this interpretation is the finding that, within the random group, correct responses averaged almost the same for those pairs with the "closer to negative" word correct (74.13) as for those with the "closer to positive" word correct (74.37). Further, supplementary analyses were conducted in order to rule out the possibility that such a response bias may have contributed to the comparisons between the random and rule-determined conditions. The random-group lists were divided into two subsets—those eight pairs in which the "closer to positive" word was correct, and those eight pairs in which the "closer to negative" word was correct—and the number of correct responses in each subset was multiplied by two. Then separate analyses of variance were used to compare each of the four rule-determined conditions with the appropriate subset of random-group pairs; e.g., the positive group was compared with the "closer to positive" set. The results of these analyses mirrored the results of Dunnett's test reported above.

It might also be pointed out that the obtained asymmetry between the positive and negative rules could not be accounted for by a failure of the three sets of words to mark off two psychologically equal intervals along the evaluative dimension. That is, assume that, either because of deficiencies

in the sample of words selected or because of some peculiarity of the evaluative dimension per se, it was harder to differentiate between neutral and negative words than between neutral and positive words. This should make it harder to apply both rules to the neutral–bad than to the neutral-good pairs. It should not make it harder to apply the negative rule than to apply the positive rule, however, since the specific pairs and, hence, the specific intrapair intervals would necessarily be the same regardless of rule.

It appears, therefore, that the negative rule was harder to discover and/or harder to utilize than the positive rule, and that this outcome was neither an artifact of unequal intervals nor the consequence of a general bias toward positive and against negative words.

Turning to the type of pair effect, performance for the five groups combined was significantly better on the neutral–good pairs than on the neutral–bad. Although the Rule × Type of Pair interaction failed to reach conventional significance levels ($.05 < p < .10$), Table 1 shows that the type of pair effect was not manifest in all groups, but was concentrated in the negative and instructed–negative conditions. This marginal interaction suggests that the superiority of the neutral–good pairs should not be interpreted simply in terms of psychologically unequal intrapair intervals. Rather, the finding of better performance on neutral–good than on neutral–bad pairs in the two negative-rule conditions is consistent with a frequency bias, since frequency would favor pairs with the neutral word correct, and under the negative rule those would be the neutral–good pairs. Further evidence for a frequency bias is found in the random group. There, frequency would not affect the comparison between neutral–good and neutral–bad pairs, but would favor those pairs with the neutral word correct as opposed to those with the extreme word correct. The figures in Table 1 (76.62 vs. 71.88) are consistent with this expectation. In the positive and instructed–positive groups, however, any frequency bias would work to the advantage of the neutral–bad pairs, and there is no evidence of such a bias. It may be speculated that the tendency to choose the more frequently occurring word played a substantial role in those conditions where learning was entirely or mainly by rote, but not in those conditions where learning was largely by rule.

TABLE 2

MEAN CORRECT RESPONSES ON WORD LIST, BY RULE CONDITION AND AWARENESS CATEGORY OF *S*

Rule condition	"Unaware" *S*s: *N*	"Unaware" *S*s: Correct responses *M*	"Unaware" *S*s: Correct responses Range	"Aware" *S*s: *N*	"Aware" *S*s: Correct responses *M*	"Aware" *S*s: Correct responses Range
Negative	6	148.33	122–172	2	181.50	176–187
Instructed–Negative	5	152.80	133–169	3	184.67	180–189
Positive	5	174.60	163–184	3	181.33	175–186
Instructed–Positive	0	—		8	189.75	188–192

Questionnaire. The 32 *S*s in the four rule-determined conditions were labeled "Aware" or "Unaware" on the basis of their responses on the postlearning questionnaire. A total of 16 *S*s checked Statement 2 indicating that a rule seemed to apply. All these *S*s then wrote rules which, while variously phrased, may be considered correct or closely correlated with the correct rule. These 16 *S*s were labeled "Aware." The 9 *S*s who checked Statement 1 to indicate that no rule seemed to apply were labeled "Unaware." The remaining 7 *S*s checked Statement 3 to indicate that a rule seemed to apply to some but not all pairs. Since none of them wrote a rule which could be considered correct or partly correct, these 7 *S*s were also labeled "Unaware." (There was essentially no difference in learning between these two subgroups within the Unaware category.)

Table 2 shows the number of *S*s labeled Unaware and Aware according to the criteria specified above. Also shown is the number of correct responses (out of a maximum possible of 192) for these two classes. All *S*s in the instructed–positive condition fell in the Aware category and, in view of their near-perfect performance, they probably discovered the positive rule on the first

trial. In the other three rule-determined conditions, only a minority of *S*s had discovered the correct rule by the end of the 12 learning trials. The range columns indicate that, within the two negative-rule groups, there was no performance overlap between Aware and Unaware *S*s. Within the positive condition, VD performance was not so strongly related to verbalization of the rule.

In view of the relatively high level attained by the five positive-condition *S*s classified as Unaware (see Table 2 and Figure 1, Panel B), their performance was compared with that of the eight *S*s in the random condition, and the difference was found to be significant, $F\ (1, 11) = 13.18$, $p < .01$. It thus appears that awareness of the positive rule—defined in terms of verbalization of the rule on the questionnaire—was not a necessary condition for its effective operation. (It might be noted that the mean correct of 174.60 among the five *S*s in question reflected equally the performance of three *S*s who checked Statement 1 and two *S*s who checked Statement 3; mean correct responses in those two subgroups averaged 174.33 and 175.00, respectively.)

A surprising finding was that the instructional clue, which had a marked effect both on the questionnaire results and on VD performance under the positive rule, had virtually no effect under the negative rule. A reasonable supposition is that an instructional clue which calls attention to evaluative attributes but does not furnish a specific rule has the effect of predisposing *S*s to try the positive rule, or an approximation of it, as their first hypothesis. In the instructed–positive group, this led to immediate discovery of the correct rule and near-perfect VD performance. In the instructed–negative group, when this first hypothesis proved wrong, most *S*s apparently reverted to rote learning and, accordingly, performed at the random-group level.

In considering the applicability of the bipolar model to performance in the present task, the form in which *S*s verbalized the rules is highly relevant. Of the 16 Aware *S*s (i.e., those *S*s who stated some version of the correct rule), only 6 stated a relational rule applicable to both types of pairs (e.g., "The more positive of the two words appeared" or "Pick the less appealing word of the pair"). The remaining 10 *S*s used absolute rather than relative terms to state, in effect, a separate rule for each type of pair (e.g., "A word associated with positive emotions appeared on the right; in pairs with negative emotions, the neutral word appeared," or "Anytime one of the words in the pairs had a positive value associated with it, the neutral word is shown on the right. A negative value appears on the right"). Assuming that *S*'s statement of the rule on the questionnaire corresponded to his formulation of it during learning, it appears that most *S*s who discovered a workable rule did not utilize the idea of an underlying good–neutral–bad continuum, since they failed to conceptualize the choice of a neutral word in some pairs and an extreme word in other pairs as following from the same relational rule. (If the relevant dimension were size instead of evaluation, an analogous outcome would be for most *S*s to formulate a rule such as, "When small is paired with medium, pick small; when large is paired with medium, pick medium" instead of the simpler relational rule "Pick the smaller one.")

Experiment II

A distinction can be made between the discovery of a rule and its utilization. The learning data combined with the questionnaire results of Experiment I indicated that the negative rule was harder to discover than the positive rule. Experiment I left indeterminate the question of whether the negative rule may also be harder to utilize than the positive rule. The purpose of Experiment II was to make a simple and direct test of the relative difficulty of applying these two rules. The task was one of judgment rather than learning.

Method

A positive group and a negative group received the same treatment except for instructions. The 16 pairs from Experiment I were presented on a memory drum at a 2-sec. rate, without feedback.

The positive group received the following instructions: "Decide which word would seem more good, or at least less bad, to most people, and call out that word. That is, of the two words, call out the one most people would say is closer to being good." The negative group received the same instructions except that the words "good" and "bad" were interchanged. (A normative rather than an idiosyncratic judgment was requested for consistency with Experiment I; i.e., in Experiment I, those *Ss* who discovered a rule would presumably have applied it on the assumption that the rule had a normative basis.) Both pairings from Experiment I were used. Within each pairing, there were two versions which were reversed with respect to intrapair top-bottom position. Lists were balanced over groups. The *Ss* were run individually. Each *S* received four trials, with a switch from one rule to the other after Trial 2. Data are reported for Trial 1 only, however, in view of the small number of errors made on Trials 2–4 and the difficulty of evaluating transfer effects.

The *Ss* were 32 paid summer session students, 16 men and 16 women, none of whom had participated in Experiment I or in the original rating of the words to be used. Sixteen *Ss*, 8 men and 8 women, were assigned to each of the two groups, using a block randomization procedure. Five additional *Ss* were eliminated, 3 because of *E* errors, 1 because he refused to judge several pairs, and 1 because he consistently selected all neutral words as "closer to good" and all extreme words as "closer to bad."

Results and Discussion

Results are reported in terms of mean errors, an error being either a response which is the normatively "wrong" member of its pair, or a failure to respond. The number of errors made on the 16 pairs ranged from 0 to 3 in the positive group, with a mean of .94, and from 0 to 7 in the negative group, with a mean of 2.19. On the neutral–good and neutral-bad pairs, respectively, mean errors were .75 and .19 in the positive group, 1.44 and .75 in the negative group. Because of apparent skewness and heterogeneity of variance, the analysis was performed on a Freeman–Tukey square root transformation of the error data (Mosteller & Bush, 1954). A $2 \times 2 \times 2$ analysis of variance with rule, sex, and type of pair as the three variables yielded significant effects for rule, $F(1, 28) = 5.01$, $p < .05$, and for type of pair, $F(1, 28) = 8.08$, $p < .01$. No other main effects or interactions were significant.

Considering first the main effect of type of pair, performance on the neutral–good pairs was significantly worse than on the neutral–bad. This was in the opposite direction from the type of pair effect found in Experiment I. The different outcomes with respect to this variable reflect different tasks; i.e., Experiment I involved a mixture of rote VD learning, rule discovery, and rule utilization, while Experiment II involved only rule utilization. The Experiment I outcome may be attributable mainly to a frequency bias affecting those pairs learned by rote. The Experiment II result could be interpreted to mean that the psychological distance between the neutral and positive words was less than that between the neutral and negative words, but it is indeterminate whether this finding can be generalized beyond the specific pairs used. As indicated earlier, any deficiencies in the samples of words selected to represent the good, neutral, and bad sets would not bias the comparison between the positive and negative rules, but might produce a significant type of pair effect.

The critical outcome of Experiment II was the rule effect. The negative rule proved to be harder to apply than the positive rule. That is, for the same pairs presented in the same way, it was significantly harder to judge which member of the pair was closer to the negative end of the evaluative dimension than to judge which member was closer to the positive end. This result held equally for neutral-good and for neutral–bad pairs; the Rule × Type of Pair interaction yielded $F < 1$.

The combined results of Experiments I and II, therefore, indicate that the negative rule is both harder to discover and harder to apply than is the positive rule.

The superiority of the positive rule in the present study bears at least a superficial resemblance to an outcome which has been found consistently with certain reasoning problems involving evaluative comparisons. That is, linear syllogisms (three-term series problems) of the form, "A is better than B, B is better than C, which is best?" are solved more readily than logically equivalent problems of the form, "C is worse than B, B is worse than A, which is worst?" (e.g., Clark, 1969; De Soto, Lon-

don, & Handel, 1965). It is possible that there is a basic asymmetry, one of wide generality, in the ability to use relational rules based on evaluative attributes. It cannot be assumed, however, that the same cognitive operations underlie performance in these different tasks.

REFERENCES

Boucher, J., & Osgood, C. E. The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 1–8.

Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **76**, 387–404.

De Soto, C. B., London, M., & Handel, S. Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, 1965, **2**, 513–521

Di Vesta, F. J., & Ingersoll, G. M. Effect of semantic redundancy on children's identification of verbal concepts. *Journal of Experimental Psychology*, 1969, **82**, 360–365.

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

Haygood, R. C. The use of semantic differential dimensions in concept learning. *Psychonomic Science*, 1966, **5**, 305–306.

Mosteller, F., & Bush, R. R Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. *Theory and method*. Reading, Mass.: Addison-Wesley, 1954.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, 76(1, Pt. 2).

Rhine, R. J. Preference for a positive evaluative response in concept learning. *Journal of Experimental Psychology*, 1965, **70**, 632–635

Rhine, R. J., Cole, D. F., & Ogilvie, L. P. Acquisition and change of an evaluative concept as a function of word frequency and type. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 55–57.

Taylor, C. L., & Haygood, R. C Effects of degree of category separation on semantic concept identification. *Journal of Experimental Psychology*, 1968, **76**, 356–359.

Wickens, D. D., & Clark, S. Osgood dimensions as an encoding class in short-term memory. *Journal of Experimental Psychology*, 1968, **78**, 580–584.

(Received June 6, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 423–430

CODING PROCESSES IN ACTIVE AND INACTIVE MEMORY [1]

GREGORY L. PETERS [2]

Ohio State University

This study examined reaction times to probes of active memory and inactive memory. Matches were based on either physical identity (e.g., the letters AA) or associational identity (e.g., the letters Aa). The results showed that inactive memory probes were processed more slowly than probes of active memory, and that physical matches were faster than associational matches only in active memory. These results were interpreted in terms of the nature of the stored representation and its retrieval. It was concluded that differences between active and inactive memory searches were due to differences in the storage location, not the nature of the stored representation.

Verbal materials have been used to study memory since Ebbinghaus' (1954) pioneer experimental work in this area in 1885. One hundred years of this type of research has promoted the notion that memory in verbal form is of much greater relative importance than other nonverbal forms (graphic, tactile, etc.). Many theories of information storage and retrieval (e.g., Atkinson & Shiffrin, 1968) have been based entirely on the registration, transfer, and storage of linguistic or verbal information.

There are, however, indications that not all storage is verbal. In an attempt to examine coding processes, Posner and Keele (1967) presented to their *S*s 2 letters which, when viewed sequentially, were judged as *same* or *different*. These judgments were based on physical identity, which occurred when 2 letters (AA) were identical physically, or associational identity, which occurred when 2 letters had only the same name (Aa). They have shown that physical *same* matches are faster than associational *same* matches. The difference (name reaction time [RT] − physical RT = 80 msec.) was interpreted as a difference in processing time for different "nodes" or levels. The assumption was made that a physical identity match is based on a comparison of codes representing only the visual information, while associational matches are based only on a comparison of verbal labels or names.

Posner and Mitchell (1967) gave 2 possible explanations for the associational–physical match time difference. First, physical matches could be made prior to associational matches, or second, both comparisons could occur simultaneously, with the associational match taking longer to complete. An alternative explanation is that in the associational match condition (Aa) one item is recoded into the format of the other, and physical matches are then made. Swanson, Johnsen, and Briggs (1972) supported the hypothesis that the memory item(s) is recoded, not the final item (probe).

In short, it seems that with the comparison of 2 successive stimuli, the identity judgments are facilitated when physical identity exists. Such facilitation lasts for at least 1.5 sec. and probably as long as 6–9 sec. or longer (Phillips & Baddeley, 1971). So, it appears that visual information is used primarily in active or short-term memory, and such memory is not entirely verbal.

A distinction that is frequently made

[1] This study was sponsored in part by the National Science Foundation through Grant GN-534.1 from the Office of Scientific Information Service to the Computer and Information Science Research Center, Ohio State University. It was based on a dissertation presented in partial fulfillment of the requirements for the PhD degree to Ohio State University. Special thanks are due to George E. Briggs who served as the author's advisor and chairman of his dissertation reading committee, and to the reviewer for his valuable comments and suggestions.

[2] Reprint requests should be sent to Gregory L. Peters, who is now at the Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio 45433.

involving storage of information is that between short- and long-term stores or active vs. inactive memory. Waugh and Norman (1965) were among the first to make such a distinction and provide empirical support. Their conclusion that the 2 systems were independent was based on differential effects of number of intervening items. Subsequent research has supported this distinction in terms of serial position recall differences (Glanzer & Cunitz, 1966; Loftus & Wickens, 1970) as well as physiological evidence (Milner, 1970). Baddeley (1966) suggested that semantic variables were most important in inactive memory, while phonemic ones were important in active memory. Shulman (1970, 1971) showed that this may be due to temporal and strategical factors, and that both variables can be effective in both systems.

Several recent studies are relevant to the question of information processing in active and inactive memory. Forrin and Morin (1969) combined the Sternberg (1966) fixed- and varied-set procedures (active and inactive memory) by presenting A, AB, or ABC (memory load [M] = 1, 2, or 3) to be retained across a block of trials (fixed set). This set is presumably held in inactive memory. Each trial, then, consisted of presentation of a varied set (M = 1, 2, or 3), followed by a probe of either active or inactive memory, or neither. This classification task is assumed to load both memory systems at the same time. They found that the RT – memory-load function for active memory probes was steeper (slower processing) than that of the inactive memory probes. However, the inactive set in Forrin and Morin's study was both highly familiar and practiced (relative to the different active sets) as well as overlapping in the formation of the memory groups. This overlap also could conceivably reduce the slope constant.

Sternberg, Kroll, and Nasto (1969) compared performance on active set (only) with performance on inactive set (only). There was less difference in practice of inactive and active probes than in Forrin and Morin's (1969) study and, because the same items (digits) were used in both tasks, there was no differential familiarity. Sternberg et al. found that the processing rate for the inactive memory probe was slower than for the active memory probes. A greater intercept for the inactive memory probes was interpreted as the time required to find the proper ensemble in inactive memory, while the increase in slope constant reflected a serial retrieval process of the inactive items into an active memory for comparison.

The present study questions this explanation of the slope difference and proposes an alternative explanation in terms of the nature of the coding used in binary classification tasks as a function of the presumed location of the stored memory set—i.e., active or inactive memory. It is predicted that visual or representational codes will be used in what is presumably active memory when verbal associates are not required for appropriate responding, while coding in what is presumably inactive memory will be based on the verbal labels even when these labels are not strictly required. These representational differences, then, might explain the processing rate differences between active and inactive memory probes. It is important to note that the distinction between active and inactive memory in this research is an operational one. A theoretical distinction, however, may find support in the results.

METHOD

To carefully examine the question of representational differences in different memory systems, acoustic similarity was manipulated in such a manner as to affect RT as a function of the presumed type of match. It was felt that acoustic similarity should alter the RT-memory-load functions only when the stimulus letters are in fact stored as associates, and these associational representations are used as the basis of comparison. That is, when physical representations are compared, acoustic similarity is not expected to be a potent factor in the comparison process.

Design. A repeated measures design was used in which 48 conditions were defined by the factorial combination of the following variables: memory load (1, 2, or 4 letters), memory system (active or inactive), type of match (physical or associational), acoustic similarity (similar or dissimilar), and type of response (match or no match). During each session 2 blocks of 84 trials were given and

the first 4 trials were excluded from analysis. Twenty of the remaining trials involved a no match of the probe and memory items and required a negative response. Of the remaining 60 trials, there were 15 trials each of the following types of match: (*a*) physical match of probe with active set, (*b*) physical match of probe with inactive set, (*c*) associational match of probe with active set, and (*d*) associational match of probe with inactive set.

The presentation of each type of material was randomized to minimize guessing. One level of memory load for the inactive set was used each session and the levels were counterbalanced so that across *S*s each level was used equally often in each ordinal position (first, second, or third) across 3 consecutive days. The inactive memory load was constant within each session to minimize confusion as to the contents of this set. All 3 levels of memory load for the active sets were used equally often in each block (10 trials/block). Given that restriction, the order of occurrence of the active memory load levels was randomized. Half of the *S*s had memory sets and probes drawn from an acoustically similar list for the first experimental sessions and a dissimilar list for the final 3 sessions. For the remaining *S*s the order was reversed.

Stimulus material. Nineteen letters of the alphabet were selected as stimuli and divided into 2 (overlapping) sets of 12 from which both memory items and probes were drawn. One set was composed of the following 12 acoustically similar letters: B, C, D, E, F, G, P, S, T, V, X, and Z. The second set was an acoustically dissimilar list comprised of the following 12 letters: A, C, D, H, K, M, Q, S, U, X, Y, and Z. Indices of acoustic confusion (Cónrad, 1964) were calculated separately for each set of letters, revealing much less acoustic similarity for the acoustically dissimilar list than for the acoustically similar list. The indices of visual confusion within each list (Townsend, 1970) were nearly identical.

The presentation of the memory sets was visual and simultaneous. For $M = 1$, the letter was in the upper left corner of a 2×2 cell matrix. The remaining cells were unfilled. For $M = 2$, the upper 2 cells contained the letters, and all 4 cells contained letters for the $M = 4$ condition. Both upper- and lowercase could be represented in one memory set. The occurrence was randomized; but, across a block of trials, both cases were represented equally often in the active memory sets.

The probe letters that could be in either case appeared in the center of *S*'s viewing screen and were selected from the same subset as the memory items. The probes were selected such that across the final 80 trials in each block there were 60 match and 20 no-match trials with an equal number of (*a*) upper- and lowercase probes (40 each), (*b*) associational matches and physical matches (30 each), (*c*) match with active set item and match with inactive set item (30 each), and (*d*) active set match probes of $M = 1$, 2, and 4 (10 each).

Procedure. In the initial session, instructions were given and *S*s had 1 block of 84 trials. All 3 levels of memory load were used for the active sets, but only $M = 4$ was used for the inactive set. The purpose of this was to develop a tendency for *S*s not to group or combine the memory sets into one set which is then rehearsed as a unitary active set. This, of course, is most difficult to do when the largest sets are used.

At the end of each session, *S* was shown his inactive set for the next session and asked to memorize it. At the beginning of the next session, *S* was asked to record his memory set to assure its storage in inactive memory. If the set or any part could not be recalled, he was again shown the entire set and given 5 min. to memorize it. Then, at the beginning of each block of trials, this set was displayed again and the regular sequence began. Following the inactive set, an active set was displayed for 3 sec. followed by a 1-sec. interstimulus interval. The probe item was then displayed and terminated after 2 sec. The time from the onset of the probe item to the initiation of *S*'s response was recorded as the primary dependent variable. The response was made by pressing a microswitch with the left or right index finger. For half of the *S*s, the left finger indicated a match while the right indicated no match. The reverse was true for the other *S*s. Following the probe there was a 3-sec. intertrial interval prior to the presentation of the next active set. Thus, the total duration of one trial was 10 sec. This sequence was repeated until the end of the block of trials. After a brief rest, the inactive set was again displayed, and the sequence was repeated with different active memory sets.

Subjects. The 24 right-handed *S*s were volunteers responding to an ad in the campus newspaper. Each was paid \$1.25 for each session. There was 1 practice session, 6 experimental sessions, and 1 debriefing session on 8 successive work days.

Apparatus. The *S* sat in a dimly lighted, sound-damped room before a response panel and a ground-glass screen. On the response panel, which was at tabletop level, were 2 buttons marked "match" and "no-match." The marking, of course, corresponded to the particular finger assignment of each *S*. The stimulus materials were rear projected at eye level on the screen and measured 20×20 mm. for the 2×2 array. Each letter was approximately 10 mm. high and was viewed at 60 cm. (less than 1° visual angle). A paper-punch-tape unit controlled the 2 Kodak Carousel RA950 projectors which presented the proper memory sets and probe items. An electronic clock (accurate to 1 msec.) measured *S*'s RT to the probe item. An Iconix timing system was used to control the several intervals and stimulus durations.

RESULTS

Each *S* contributed 10 positive RTs (excepting errors and machine failures) to

TABLE 1

AVERAGE CORRECT REACTION TIMES (IN SEC.) AND RELATIVE ERRORS (IN PERCENTAGES)

Type	Memory load					
	1 letter		2 letters		4 letters	
	RT	RE	RT	RE	RT	RE
Inactive						
Physical active						
1	439	.41	454	.62	469	.83
2	465	.80	497	.62	510	1.26
4	535	2.78	554	3.64	566	8.63
Associational active						
1	469	1.04	474	.83	489	1.46
2	497	1.04	516	1.73	530	2.08
4	558	10.29	570	3.78	580	2.50
Active						
Physical inactive						
1	480	.20	490	.20	547	2.55
2	548	.83	562	1.69	608	3.13
4	608	4.00	639	6.49	656	3.45
Associational inactive						
1	476	.41	500	.20	547	1.93
2	557	1.26	577	1.93	622	1.46
4	626	2.74	639	5.57	682	3.57

Note. Abbreviations: RT = reaction time; RE = relative errors

each of the cells in Table 1 for each similarity condition (acoustic similarity or dissimilarity). Medians of these scores were taken, while all further measures of central tendency were derived by use of arithmetic means. Table 1 shows the average correct RTs and percent errors for each condition, collapsed across both *S*s and similarity conditions.

Error analyses. The attempt to maintain a low error rate was shown to be successful by the 2.31% average error rate for match conditions and 8.74% for no-match conditions. The error rate data provide little additional information beyond that shown below by an analysis of RTs. It is important, however, to note the lack of significance of the acoustic similarity variable in terms of errors, $F\ (1, 23) = 2.96$, $p > .10$. This suggests that the comparison of RTs for the levels of acoustic similarity is valid.

Reaction tim [illegible] A detailed analysis of no-mat [illegible] is difficult because such resp [illegible] not be categorized as arising from [illegible] either the active or inactive men [illegible] It was expected that no-match [illegible] would be slower than match res [illegible] due only to the lower probabili [illegible] rence, and this expectation w [illegible] orted (no-match grand $M = 720$ [illegible] match grand $M = 543$ msec.).

An examinat [illegible] tch responses shows that the [illegible] in effects were significant at [illegible] mory system, $F\ (1, 23) = 4$ [illegible] 083; type of match, $F\ (1, 2$ [illegible] $= .077$; and memory load, [illegible] 13, $\omega^2 = .272$. Also, 2 first-ord [illegible] reached significance: Mem [illegible] $\times$ Type of Match, $F\ (1,$ [illegible] and Memory System $\times$ Mem [illegible] $(2, 46) = 3.63$.

The significan [illegible] main effects supports the re [illegible] merous prior studies. First, [illegible] superiority of active set prob [illegible] 9 msec.) over inactive set prob [illegible] ec.) was found earlier by Ster [illegible] al. (1969) and Forrin and Mori [illegible] 69). This distinction between active and inactive memory is supported by the fact that, in a postexperimental questionnaire, 23 of 24 *S*s reported not rehearsing the inactive set items.

Second, the main effect of type of match shows that physical matches (535 msec.) were made faster on the average than associational matches (551 msec.). Posner and Keele (1967), Posner et al. (1969), and others also found this difference. The magnitude of the difference found here (16 msec.), however, is smaller than that found by Posner and Mitchell (1967). This is due to the use here of the mixed-list procedure (Posner et al., 1969), inclusion of some analogue matches, the 1-sec. interstimulus interval, and the inclusion of associational–physical differences from both active and inactive memory set probes.

Finally, the characteristic increase in RT with an increase in memory load (Sternberg, 1966) is evidenced in the significant main effect of memory load.

The significant Memory System × Memory Load interaction, illustrated in Figure 1, suggests that the slopes of the 2 functions are not parallel. The best-fitting linear equations are

RT = .440 + .068 (H_c) Inactive Memory

and

RT = .417 + .047 (H_c) Active Memory.

It is important to note that the H_c scale of the number of items in memory is an equal-interval log transform scale. The H_c metric is a Shannon statement (log) of central processing uncertainty which is determined primarily by memory load (see Briggs et al., 1971). Such a scale is not without precedent (Briggs et al., 1971; Nickerson & Feehrer, 1964) and more important, the data overall are better fitted in this manner than by a linear scale.

The prediction that the type of information used in the comparison process will be a function of the particular memory system involved (active vs. inactive memory) is supported by the significance of the Type of Match × Memory System interaction. This interaction is plotted as a function of memory load in Figure 2, and indicates that the main effect of type of match was due to the difference between associational and physical matches in active memory. For this reason the physical and associative match functions for inactive memory,

RT = .438 + .066 (H_c) Physical

and

RT = .443 + .069 (H_c) Associational,

are plotted as a single regression equation in Figure 2.

Thus, the data in Figure 2 support the major prediction that the information used to match stimuli in this task is a function of the memory system involved. Further support comes from the contribution of the significant type of match main effect and from the Type of Match × Memory System interaction. These effects show that physical matches are faster than associational matches, but that this difference occurs in the active memory system. Thus, only associational matches are involved in the inactive memory system, while both types of matches occur in active memory.

The fact that neither Type of Match × Memory Load nor any other higher order interactions involving both of these variables was significant indicates that the difference between associational and physical matches in the active memory set is constant across memory load and thus the functions are plotted as best-fitted parallel lines in Figure 2.

Neither the main effect of acoustic similarity, F (1, 23) = .34, nor any of its

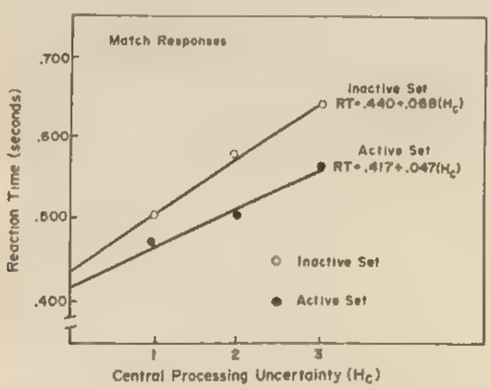


FIGURE 1. Average "match" reaction times (RTs) as a function of central processing uncertainty (H_c) with memory system as the parameter.

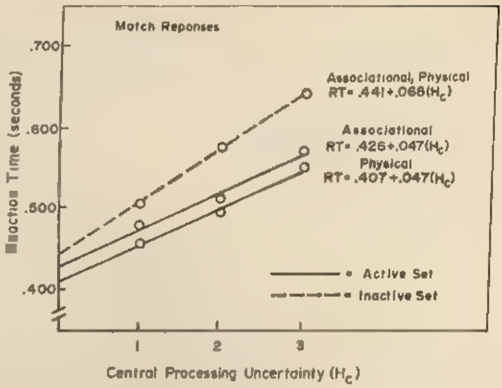


FIGURE 2. Average "match" reaction times (RTs) as a function of central processing uncertainty (H_c) and memory system with type of match as the parameter.

interactive patterns approached significance. It was expected that only the associational matches would be affected by acoustic confusability, and such an effect would serve to converge on the representational differences presumably existing in the different memory systems. If it is assumed that acoustically mediated associations are stored in inactive memory, then it follows that these associations would effect RTs to associational matches in active memory more so than physical matches in active memory. Table 2 revealed that this did not occur. In fact, these data show that that which mediates matches in inactive memory must be operating at least as much in active memory physical matches as in active memory associational matches.

Discussion

Previous studies have loaded what is presumably active and inactive memory in character classification tasks (Forrin & Morin, 1969; Sternberg et al., 1969). It was suggested that differences between these studies might be artifacts of control procedures, such as practice, familiarity, overlap of memory sets, etc. The present study sought to achieve proper control over these variables and determine other and possibly more meaningful explanations for differences in active and inactive memory probe RTs. The following major predictions were made: (*a*) inactive probe RTs are slower than active probe RTs, (*b*) the rate of processing in inactive memory is slower than in active memory, (*c*) both associational and physical matches are made in active memory, and (*d*) only associational matches are made in inactive memory. Support for these predictions is provided in Figure 2. These data suggest that the information stored in active memory is such that both associational matches and physical matches are possible. This in turn implies that visual information, which is more or less isomorphic with the stimulus, is represented in what is presumed to be active memory and is more efficient for use in matching than the associational information. Further, the data suggest that visual information becomes relatively less efficient than the associational information in inactive memory or it must be retrieved in a recoding operation. Thus, *Ss* are able to select and to use visual information when searchi[...] memory in a recognition task; but, [...]tive memory is searched, they are l[...] to use such information.

Three prominent feat[...]igure 2 are most important: (*a*) t[...]nt associational-physical match [...] in active memory only, (*b*) the slo[...]t difference between active and inac[...]y, and (*c*) the intercept difference [...]ctive and inactive memory. Th[...]ce between associational and physi[...]es can be considered to be an es[...] additional processing time require[...] an associational match above [...] a physical match. This additional [...]g will be called recoding and cert[...]haracteristics are examined in [...]il below. Swanson et al. (1972) fo[...]th active memory probes, the diff[...]een associational and physical mat[...]anifested in the slope of the RT – m[...] function, not in the intercept as n[...] There is, however, a fundamental [...]e in the constitution of the memo[...]ween the 2 experiments. The Sw[...] memory sets were homogeneous (i[...]ems were of one format) allowing se[...]ing of the memory set items into th[...]at as the probe (forms vs. digit nam[...] the slope

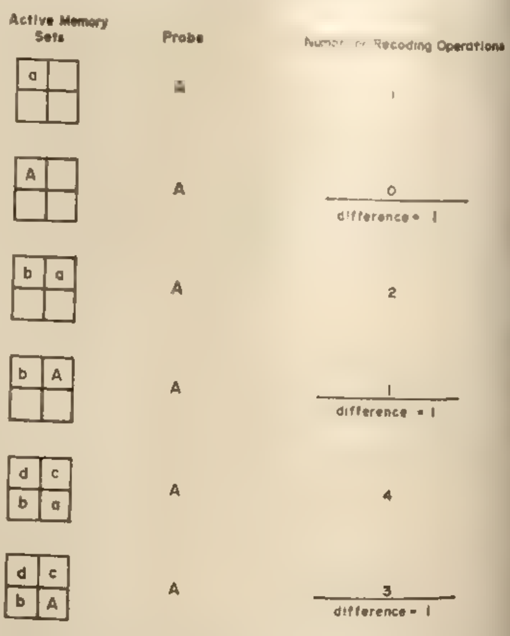


FIGURE 3. Diagram of recoding operations to explain constant difference between physical and associational matches in active memory.

effect. Here, the memory sets were heterogeneous, i.e., each set could contain both upper- and lowercase letters. Thus, if the memory items are recoded (as Swanson et al. suggested) this must be done on an individual basis. That is, all memory items cannot automatically be recoded given the probe item; each recoding must be based on a comparison of each memory letter with the probe.

Figure 3 diagrams several possible physical and associational match conditions. Its purpose is to illustrate associational–physical match differences in terms of the number of recoding operations. Several points must be considered in examining this figure. First, the matching memory item is shown always in the "final" position. There is neither evidence in this experiment that a serial scan exists, making this the final position, nor that the scan is exhaustive. Also, appearance of the probe in other positions changes the numerical values in the figure but not the difference or conclusions. In sum, the figure involves several noncritical, simplifying assumptions. Second, by definition, associational matches require recoding. This recoding occurs whenever a memory item and the probe item do not match in terms of physical identity. That is, both a (memory item) – A (probe) and D (memory item) – A (probe) require recoding. Thus, for example, in Figure 3 when ab is presented as the memory set and A is the probe, 2 recoding operations are necessary. The third important point is that the difference between associational and physical matches is represented as recoding operations and the number of such operations is constant (difference = 1) regardless of the number of items in the memory set. It is proposed that this constancy yields the intercept difference between associational and physical matches and that the magnitude of the difference (19 msec.) is due to the necessity of one additional recoding for the associational match situation.

The slope difference between active and inactive memory functions is also of concern. First, processing of items presumably stored in inactive memory is slower than processing of active memory probes. This precludes Forrin and Morin's (1969) concept of a match in inactive memory which does not involve search and is independent of the number of items in that memory system. A more reasonable explanation is that items stored in inactive memory must be retrieved prior to comparison with the probe. A parallel retrieval would result in an intercept-only difference between active and inactive memory functions. The slope difference does suggest a serial retrieval of the inactive memory items. The notion that this difference is due to comparison of different codes is rejected because the slope of the associational matches in active memory was not the same as the slope in inactive memory. This introduces the second important point. If recoding of physical into associational form is assumed to involve some interaction with inactive memory (e.g., the association between A and a, be it verbal or visual, is stored in inactive memory), then it would be expected that the estimate of recoding time should be the same as the retrieval of information stored in inactive memory. Figure 2 shows that the recoding time for active memory items is approximately 20 msec. The difference in slopes between active and inactive memory functions is interpreted as an estimate of the serial retrieval time for the inactive memory items. This difference is also approximately 20 msec. Thus, recoding time and retrieval time, which have logical commonalities, are estimated to be of approximately equal duration. This is not to say that recoding and retrieval can replace one another. If this were so, then retrieval of the inactive set items would eliminate the necessity for subsequent recoding, which again, would produce equal slopes in Figure 1.

The third predominant feature in Figures 1 and 2 is the intercept difference between the active memory and inactive memory functions (23 msec.). That this is due to parallel retrieval processes was rejected above (in favor of serial retrieval). Sternberg et al. (1969) suggested that it is due to the additional time required to locate the appropriate ensemble of items in inactive memory. A third possible explanation is that this intercept difference reflects time taken to "identify" or recode the probe item into its associate form. This, however, assumes that active memory is examined first and such identification follows the examination and is prior to search of inactive memory. Such an assumption finds some support in the present study (Table 1); however, earlier it was shown that any recoding of the probe, in connection with active memory, is unlikely due to the mixing of cases in the memory set. Because the same logic applies here, the most parsimonious explanation for the difference in intercept between active and inactive memory

functions is simply the "time to locate" (Sternberg et al., 1969) the inactive set ensemble.

It was predicted that acoustic similarity would increase RTs when associational matches were made. If so, then the association would be presumed to be a verbal one based on an acoustically represented name. The total lack of a significant effect of this variable, coupled with a similar lack of effect of "nameability" on associational matching in a study by Taylor (1969), casts doubt on the hypothesis. There are 2 reasonable alternative hypotheses concerning the nature of the associational match. First, the association may be an abstract verbal code which involves the name of the item but not in acoustical form (Taylor called it a "plan"). Second, the association may be a recoded physical representation. That is, when S is given A to memorize followed by the probe a, he recodes A into a. This recoding is followed by a physical match of a with a. An appropriate manipulation of the visual similarity would address this issue.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 2. *Advances in research and theory*. New York: Academic Press, 1968.

Baddeley, A. D. Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 362-365.

Briggs, G. E., Fisher, R. L., Greenberg, S. N., Lyons, J. J., Peters, G. L., & Shinar, D. *Multi-task time-sharing requirements*. (Tech. Rep. AMRL-TR-71-105). Wright-Patterson Air Force Base, O.: Aerospace Medical Research Laboratory, 1971.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75-84.

Ebbinghaus, H. Über das Gedächtnis. In R. S. Woodworth & H. Schlosberg (Eds.), *Experimental psychology*. New York: Holt, Rinehart & Winston, 1954.

Forrin, B., & Morin, R. E. Recognition times for items in short- and long-term memory. *Acta Psychologica*, 1969, **30**, 12[illegible]

Glanzer, M., & Cunitz, A[illegible] storage mechanisms in free recall. *Jo[illegible] Verbal Learning and Verbal Behavior*, 1966, 5, 351-360.

Loftus, G. R., & Wickens, [illegible] D. Effect of incentive on storage and retrieval processes. *Journal of Experimental* [illegible], 1970, **85**, 141-147.

Milner, B. Amnesia [illegible] operation on the temporal lobes. In W. Kintsch, *Learning, memory, and conceptual* [illegible] New York: Wiley, 1970.

Nickerson, R. S., & Feeh[illegible], C. E. Stimulus categorization and respo[illegible] time. *Perceptual and Motor Skills*, 1964, **18**, [illegible] 793

Phillips, W. A., & Badde[illegible], A. D. Reaction time and short-term visu[illegible] memory. *Psychonomic Science*, 1971, **22**, 7[illegible]

Posner, M. I., Boies, S. [illegible] W. L., & Taylor, R. L. Retent[illegible] and name codes of single letters. [illegible] *Psychology Monograph*, 19[illegible]

Posner, M. I., & Keele, S. [illegible] decay of visual information from a single [illegible], 1967, **158**, 135-139.

Posner, M. I., & Mitchell [illegible] chronometric analysis of classification. [illegible] *Review*, 1967, **74**, 392-409.

Shulman, H. G. Encod[illegible] retention of semantic and phonemic i[illegible] in short-term memory. *Journal* [illegible] *Learning and Verbal Behavior*, 1970, **9**, 49[illegible]

Shulman, H. G. Similarity e[illegible] in short-term memory. *Psychological Bulletin*, 1971, **75**, 399-415.

Sternberg, S. High speed scanning in human memory. *Science*, 1966, **153**, 652-654.

Sternberg, S., Kroll, R. L., & Nasto, B. A. Retrieval from long-term vs. active memory. Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1969.

Swanson, J. M., Johnsen, A. M., & Briggs, G. E. Recoding in a memory search task. *Journal of Experimental Psychology*, 1972, 93, 1-9.

Taylor, R. L. Comparison of short-term memory and visual sensory analysis as sources of information. *Journal of Experimental Psychology*, 1969, **81**, 515-522.

Townsend, J. T. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 1970, **9**, 40-50.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89-104.

(Received May 7, 1973)

(of Experimental Psychology
), Vol. 102, No. 3, 431–434

INTERACTION OF SIMILARITY TO WORDS OF VISUAL MASKS AND TARGETS

J. ZACHARY JACOBSON [1]

Queen's University, Kingston, Ontario, Canada

Words masked by upright random letters in a wordlike array or by unrelated words are more difficult to read than words masked by other patterns, including strings of complete letters in random orientation. Such a difference was not observed when targets were figures to be identified. This result suggests that a wordlike mask made of letters sets up a reading response that interferes with the processing of a target when the target requires reading.

It has been reported (Jacobson, 1973) that words that are backward masked by unrelated words or by zero-order pseudowords are much more difficult to read than the same words masked by a random-dot Julesz (1971) field. Letters have a higher contour density than the random pattern used in the Jacobson study (squares half as high as the letters), and it was suggested that this localized difference in "grain" produced the aforementioned result.

However, a subsequent study called the contour-density explanation into question. Masks were made by systematically printing one letter on top of another and then erasing a randomly chosen half of each of the 2 letters (see Figure 1B). Strings of such "chimeric letters" were found to be much less effective as masks of words than unrelated words used as masks.[2] It could not be argued that the chimeric-letter masks had less contour density than the words used as masks; if anything, the chimeric letters (which most *S*s said resembled Chinese characters) had a greater contour density than the ordinary letters that comprised the words.

In an attempt to find out more about this puzzling result, a study was done to compare the masking of words by a number of stimuli purposely varied in their similarity to printed English. These masks were: Julesz fields, randomly oriented pieces of letters, randomly oriented whole letters, upright whole random letters (zero-order pseudowords), unrelated words, and words that were associates of the target words. The last condition was included as an internal check for the validity of the study. Words masked by associates are more readily reported than words masked

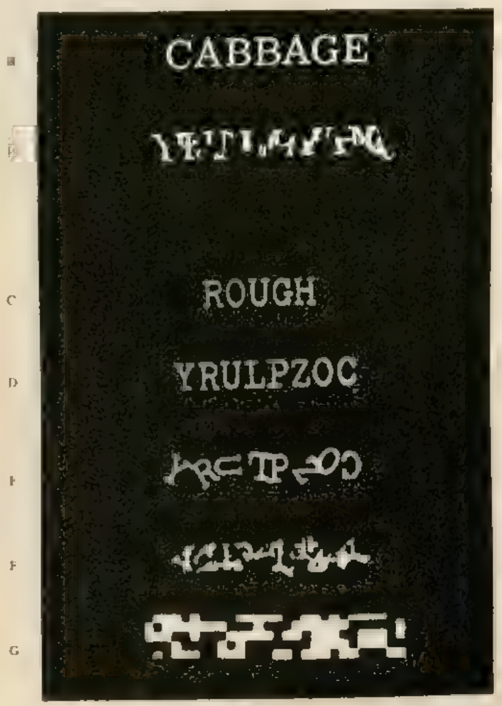


FIGURE 1. Stimuli used in a preliminary study and in Experiment I. (Rows labeled A and B = comparison of word and chimeric-letter stimuli used in the preliminary study; C–G = stimuli used in Experiment I: C = word, D = zero-order pseudoword, E = jumbled letters, F = letter pieces, and G = Julesz field.)

[1] Requests for reprints should be sent to J. Z. Jacobson, Psychology Department, Queen's University, Kingston, Ontario, Canada.

[2] Unpublished pilot data collected by the author.

by unrelated words (Jacobson, 1973); if this finding could not be replicated in the present study, the remainder of the data would be suspect. Furthermore, an attempt was being made to continuously vary the similarity of the mask to the target; including associate masks in the study formed a logical extension of that continuum in the direction of increased similarity.

Experiment I

Method. Twenty-five introductory psychology students (20 women and 5 men) served as *S*s.

Letter, word, and letter-piece stimuli were made by photographing typewritten materials with black-and-white film and mounting the negatives as slides for the projection tachistoscope. Thus, *S* saw white capital letters or pieces of letters on a black background. Sixteen associated word pairs were chosen from Arthur's (1969)[3] list of free associates (Figure 1C); unrelated words were the same words mispaired. Zero-order pseudowords (wordlike arrangements of randomly chosen letters) were also used as masks (Figure 1D). Randomly chosen letters in random orientation but arrayed on a line ("jumbled letters") formed another class of mask (Figure 1E). Pieces of letters in random order arranged on a line were yet another sort of mask (Figure 1F).

Strips of Julesz fields from Gregory (1966) were photographed to form the random-dot masks (Figure 1G). Targets for the nonword masks were chosen randomly from the same set of words as targets for word masks.

Each mask was arranged to cover its particular target when projected. A Gerbrands 2-field projection tachistoscope was used to present the stimuli in a dimly lit room. The letters projected on the screen were 4 cm. high with a luminance of 8.0 ± .5 ftL. The background was 2.0 ± .3 ftL. There was no fixation point. The *S*s were seated 3 m. from the screen.

In a pretest procedure, each *S* was shown 1 word for 8 msec. one or very few times (it is not difficult to read unmasked words at this interval under these conditions). Then, as a demonstration of masking, *S* was shown the same word followed 2 msec. later by another word presented for 120 msec. He was asked if he saw the first (target) word, and when he said he did not, he was assured that a word was indeed shown. The masking sequence was slowed down and shown to him again to assure him of that fact. He was told that people for some reason find it difficult or impossible to read a word when it is followed so quickly by a mask. It was further explained that this was to be a study in [illegible] target word would be followed by any of sev[illegible]es of masks, and his task would be to read the target[illegible]

Following the familiarization, 24 randomly ordered trials were given. Targets were always exposed for 8 msec., masks for 120 mse[illegible]

The mask delay needed [illegible] to read a target word correctly was determined [illegible] ascending method of limits. Each trial w[illegible]d with 10 msec. elapsing from the onset of the target to the onset of the mask (Kahneman, 1968, calls this interval stimulus onset asynchrony, or SOA). If *S* could not read the target correctly, the interval was increased by 10 msec. and the seque[illegible] displayed again. To warn *S*, *E* said "ready" be[illegible] each sequence. The *S*s used this warning time [illegible] fixate that part of the screen where the stimuli w[illegible] presented. The procedure was repeated, inc[illegible] interval by 10 msec. each time, until th[illegible] read correctly. The delay of the mask [illegible] *S* read the target word correctly wa[illegible] the dependent variable.

Results. An anal[illegible] variance was conducted on the da[illegible] *S*s became slightly more facile w[illegible] 4 repetitions of each mask condition, [illegible] a mean SOA of 43 msec. on the fi[illegible]etition of each mask condition to 36 [illegible] on the fourth, $F(3, 72) = 3.64$, $p <$ [illegible] There was no apparent interaction between mask condition and repetition. Mean SOAs when the targets were read under the various conditions were, in order of less to more similarity between mask and target: Julesz-field mask, 36 msec.; pieces of letters, 32 msec.; jumbled-letter mask, 36 msec.; 0-order letter sequences, 53 msec.; unrelated words, 48 msec.; and associated words, 33 msec. (See Figure 2).

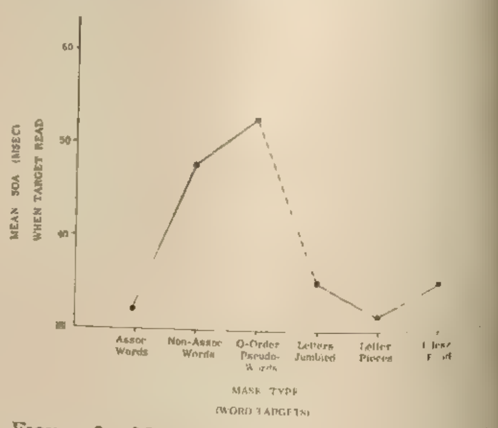


FIGURE 2. Mean stimulus onset asynchrony (SOA) when the target was read correctly: Experiment I.

[3] A. Z. Arthur. Queen's norms for responses to 100 words from the Kent-Rosanoff Word Association Test. (National Research Council of Canada Project APA 0269). Unpublished manuscript, Queen's University, Kingston, Ontario, Canada, 1969.

Individual comparisons with regard to mask treatments were done on the data. They showed that there were 2 separate kinds of masks: those that appeared like or actually were words, and those that did not appear like words, regardless of whether they were made of letters or not. Targets masked by unrelated words and by zero-order pseudowords were more difficult to read than those masked by jumbled letters, letter pieces, and Julesz fields, F (1, 24) = 60.0, $p < 10^{-5}$. There were no reliable differences among words masked by jumbled letters, letter pieces, or random dots. There was no reliable difference between words masked by unrelated words and those masked by zero-order pseudowords. There was a reliable difference between the associated and nonassociated mask conditions, F (1, 24) = 16.04, $p <$ .0008. This comparison is not orthogonal with the rest and is only presented to demonstrate that a previously noted effect was also present for these *S*s.

Discussion. Again, differences in grain cannot explain the results. Contour density was controlled in the jumbled-letter and letter-piece masks by, in the first case, centering each randomly oriented letter physically where it would have appeared if it had been a zero-order pseudoword. A similar procedure was used for the letter pieces except that 2 letters were centered in each position and a random half of each was erased. Thus, contour density was held constant across word, zero-order, jumbled-letter, and letter-piece masks.

The crucial variable seemed to be the arrangement of upright letters in a wordlike array, regardless of whether the letters actually formed a word. An after the fact explanation of this result might be that a reading set is triggered by a mask that is a word or by letters arrayed as a word, and that reading or attempted reading of the mask interferes with the reading of an unrelated target.

If this were true, the relative differences in interference with the target should be limited to reading the target word (or identifying its letters, perhaps), but not to recognizing some other property of it. To test this, another study was done in which the same masks were used, but the targets were rows of shapes for *S*s to identify. The sharp distinction between zero-order masks and the rest of the nonword masks might not occur under the foregoing suggestion.

Experiment II

Method. Sixteen students in introductory psychology, 12 women and 4 men, served as *S*s.

New targets were made of rows of squares, triangles, circles, diamonds, or arrowheads the same size as letters and arrayed in a line; each target contained only 1 sort of shape. These were masked by the same zero-order pseudowords, jumbled letters, letter pieces, and Julesz fields used in the last study. As before, trials in which words were masked by related and unrelated words were included as an internal check of the results.

Again, each *S* was presented each sort of trial 4 times. The procedure was the same as that used in Experiment I.

Results. As before, *S*s improved steadily from the first trial in each condition to the last. Mean SOA was 70 msec. on the first trial and 54 msec. on the last, F (3, 45) = 7.35, $p <$ .001. There was no interaction of repetition and mask type.

Mean SOAs when the shapes were correctly identified were: Julesz-field mask, 47 msec.; letter-piece mask, 51 msec.; jumbled letter mask, 59 msec.; and zero-order mask, 55 msec. There was no apparent difference between zero-order pseudowords on the one hand and jumbled letters, letter pieces, and Julesz fields on the other, F (1, 15) = .5342, p = .52. As before, associated targets were easier than unassociated targets, 42 msec. vs. 79 msec., F (1, 15) = 30.42, p = .0002.

There was a reliable tendency for shapes masked by letter pieces to be more readily reported than those masked by jumbled letters, F (1, 15) = 11.5, $p <$ .004.

Discussion. The *S*s experienced more difficulty identifying word targets in the second study than in the first. It is likely that this was due to the uncertainty present in the second but not in the first study as to whether a particular target was a word or a row of shapes.

It is obviously not possible to state from the observed absence of a difference in identifying shapes when masked by zero-order pseudowords relative to all other masks that no such difference exists. Nevertheless, it seems reasonable to say that such a difference, if it does exist, is much less marked than when the targets are words. This tends to confirm the

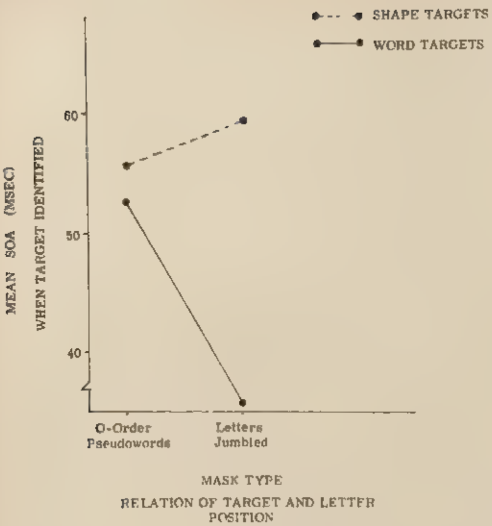


FIGURE 3. Relation of identification of words and shapes when masked by zero-order pseudowords and jumbled letters. (SOA = stimulus onset asynchrony.)

hypothesis that the wordlike arrangement of the mask sets up a process or response that interferes with the identification of a target when that target requires reading, but not otherwise.

There is no inconsistency between this explanation and the fact that words masked by associates were more readily reported than words masked by nonassociates and by zero-order pseudowords. This result cannot be ascribed to guessing, and the explanation that has been suggested is that it is a perceptual effect involving memory traces of the target and mask words (Jacobson, 1973). The memory traces of 2 associated words involve large networks that include many of the same circuits or units. The sets of circuits ("lattices") representing 2 unassociated words in memory would not have as high a degree of mutual inclusion among their elements. Thus, while masking a word by another word interferes with the processing of the target, the activity set up by the mask—if it is an associate of the target—tends to partially restimulate the target's trace. A word masked by an associate, then, needs less of an input of energy containing the target information before it can be identified than the same target word would need if it were masked by an unassociated word.

In the present study, this facilitation by association occurred and effectively countered the interference created by the wordlike appearance of the associated mask. As a result (when the target is a word) there are 2 series of masking stimuli—a wordlike series in which association is an effective aid to recognition, and a nonwordlike series of masks in which word association obviously cannot operate.

Turvey (1973) has suggested that there are 2 kinds of visual masking, one taking place peripherally and one centrally, with markedly different properties. Except for the Julesz fields, all of the masking stimuli used in the present study seem to be the sort that Turvey found to act centrally as well as peripherally. The interaction reported here between mask and target material with regard to their linguistic content (see Figure 3) indicates that, if Turvey is correct, there must be different processes acting even within the central sort of masking. Of course, this assumes that the difference between zero-order and jumbled-letter masks that was observed in Experiment I was effected centrally and not peripherally, but it is difficult to imagine how that could be otherwise.

REFERENCES

GREGORY, R. L. *Eye and brain.* Toronto: McGraw-Hall, 1966.

JACOBSON, J. Z. Effects of association upon masking and reading latency. *Canadian Journal of Psychology*, 1973, **27**, 58-69.

JULESZ, B. *Foundations of Cyclopean perception.* Chicago: University of Chicago Press, 1971.

KAHNEMAN, D. Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, 1968, **70**, 404-425.

TURVEY, M. T. On peripheral and central processes in vision: Inferences from an information processing analysis of masking with patterned stimuli. *Psychological Review*, 1973, **80**, 1-52.

(Received May 9, 1973)

..al of Experimental Psychology
'74, Vol. 102, No. 3, 435–438

ORTHOGRAPHIC DISTINCTIVENESS OF CONSONANTS AND RECOGNITION LEARNING

DONALD H. KAUSLER[1] AND EDWARD J. PAVUR, JR.

University of Missouri, Columbia

The effects of varying the distinctiveness of uppercase and lowercase forms of consonants were determined for both single-item (i.e., individual study, or target, items that are later tested for recognition with paired distractors) and multiple-item (i.e., paired study items, one of which is *right*, that are also tested in pairs for recognition of *right* items) recognition learning. Both tasks revealed a greater error rate for consonants of low distinctiveness than for consonants of high distinctiveness.

A feature-analytic hypothesis of single-item recognition learning (e.g., Anisfeld & Knapp, 1968) attributes the false-recognition effect (Underwood, 1965) to a feature-tagging principle. Briefly, sensory and/or semantic features of individually presented study items are presumed to be tagged in some manner during their initial exposure. Recognition of later test items as *old* is then based on their possession of these previously tagged features. Accordingly, the false recognition of *new* items as *old* is likely to be greater when they share sensory and/or semantic features with prior study items than when they are unrelated to prior study items.

A seemingly potent source of sensory features for mediating the false-recognition effect is that of orthography. In a recent experiment by Raser (1972), a number of study items were homophones that were later paired for testing purposes with their homophonic counterparts as distractors. The pairs were either high or low in orthographic distinctiveness (e.g., CLAWS–CLAUSE and FOUL–FOWL, respectively). Also included in the experiment were nonhomophonic study items that were later paired with distractors of either high or low orthographic distinctiveness (e.g., CLOUD–CLEAVE and SOUR–SLUR, respectively). The false-recognition rate (i.e., selecting new intrapair members as *old* study items) was contrasted for test pairs of varying disinctiveness. Sharing of tagged orthographic features should, of course, be greater when the *old–new* members of test pairs are low in distinctiveness than when they are high. In agreement with this hypothesis, Raser found that orthography had a significant effect on the rate of false recognitions. In fact, the effect was more pronounced than that found for shared phonemic features.

Similar findings have been reported for the multiple-item, or verbal discrimination, recognition learning task. Kausler (1973b) discovered slower acquisition for paired homophones (i.e., one member of each pair functioned as the *right* item, the other as the *wrong* item) of low orthographic distinctiveness than for paired homophones of high orthographic distinctiveness. Schulz and Lovelace (1972) found a parallel effect when either homophones or nonhomophones of high and low orthographic distinctiveness functioned as interpair *right* and *wrong* items. These results support the general hypothesis that multiple-item recognition learning, like single-item recognition learning, is mediated by a feature-analytic–feature-tagging process (Kausler, 1973a, 1973b, in press).

The present experiments provided a further test of the effects of orthographic distinctiveness on recognition learning. Specifically, the test involved distinctive features at the level of graphemic content, rather than distinctive features at the more molar level of distinctive orthography (i.e., letter content, word length, etc.—as is the case when words serve as items). The

[1] Requests for reprints should be sent to Donald H. Kausler, Department of Psychology, University of Missouri, 209 McAlester Hall, Columbia, Missouri 65201.

present study items were consonants scaled for the distinctiveness of their upper- and lowercase forms. Upper- and lowercase forms of the same letter have identical phonemic features, and, presumably, highly overlapping semantic features as well. Consequently, when one form of each letter is a study item and the other form a paired test distractor, intrapair discriminations are likely to be accomplished through whatever distinctive graphemic features were tagged during the original study phase. Since consonants vary widely in the distinctiveness of their upper- and lowercase forms (e.g., G, g vs. Z, z), their hit rates in single-item recognition learning should reflect this variation. A similar effect should occur when the upper- and lowercase forms function as intrapair *right–wrong* (or *wrong–right*) members of a multiple-item recognition learning task. Experiments I and II provided tests of this hypothesis for the single-item and multiple-item tasks, respectively.

Experiment I

Method

Subjects and design. Forty students in general psychology classes at the University of Missouri, all of whom were naive with respect to prior participation in recognition learning experiments, served as *S*s. They were assigned alternately to 2 groups ($n = 20$) that differed only in the instructions (standard vs. experimental) given prior to the first study phase. Distinctiveness was a within-groups variable, with half of the study list composed of consonants previously rated (by a separate group) as being high in upper- and lowercase distinctiveness and the other half of consonants previously rated as being low. Thus, the design was a 2×2 factorial, with one between-groups and one within-groups variable.

Materials and lists. Upper- and lowercase forms of the 20 consonants had been rated earlier by 112 *S*s (general psychology students) for their physical distinctiveness on a 9-point scale (1 = very distinctive, 9 = very similar). Mean ratings ranged from 3.16 (G vs. g) to 8.45 (Z vs. z). For purposes of this study, the consonants were divided at the median into a high distinctiveness (HD) set (G, Q, D, R, N, H, B, L, T, and M—in order of decreasing distinctiveness) and a low distinctiveness (LD) set (Z, V, P, X, W, S, C, K, J, and F—in order of increasing distinctiveness). The uppercase forms served as study items for 5 HD and 5 LD consonants; the lowercase forms served as study items for the remaining 10 consonants. For greater generalization of results, 2 variants of the study list were employed. In the first, the uppercase form of the odd-numbered consonants (as ranked from most to least distinctive) and the lowercase form of the even-numbered consonants were study items. In the second variant, the lowercase form of the odd-numbered consonants and the uppercase form of the even-numbered consonants became the study items. The *S*s in the 2 instructional groups were assigned alternately to the 2 list variants. In addition, 2 different random orders for presenting the study items were employed per trial in each list variant. Neither list variation nor order of presentation were later found to influence hit rate. Consequently, these factors were pooled for the analyses reported later. Study slides and subsequent test items were typed on the same IBM Selectric typewriter.

List length was extended by including 30 2-digit numbers as additional study items. The first 5 and the last 5 study items were always numbers; the remaining 40 items alternated between consonants and numbers.

Procedure. All *S*s were fully informed as to the composition of the study list. That is, they knew that uppercase consonants, lowercase consonants, and numbers would be exposed, and that they were to remember the items for a subsequent memory test. They were further informed that uppercase and lowercase consonants would always appear on the right and left sides of the screen, respectively. This procedure was followed in order to assure perception of the appropriate case form of the less distinctive consonants. The standard instructional group received only the above information. The experimental instructional group received additional information suggesting that a useful device for memorizing the consonants would be to form an image of each consonant that emphasized its distinctive features. It was anticipated that these instructions would reduce the disparity in hit rates attributable to graphic distinctiveness.

There were 2 study–test trials. On the study phase of each trial the individual items were exposed, via a Kodak Carousel projector, at a 2-sec. rate, with a .9-sec. interitem interval. Test items were presented in a test booklet, with a separate page for each trial. Each *old* consonant item was paired with its upper- or lowercase counterpart as a distractor, and each *old* number with a *new* number as a distractor. Correct items appeared equally often on the right and on the left of the test pairs. Two different random orders of pairs were used per trial with each of the list variants described earlier. Order was again found to be unrelated to hit rate, and the different orders were pooled for the main analyses. For each of the 50 pairs, *S* encircled the *old* item, and rated the confidence of his judgment on a 5-point scale. The test phases were self-paced.

Results and Discussion

Means and standard deviations for total errors on the 2 trials are given in Table 1.

The main effect for distinctiveness was statistically significant, F (1, 38) = 4.75, $p < .05$. As hypothesized, the error rate was greater for LD consonants than for HD consonants. Neither the main effect for instructions nor the Instructions × Distinctiveness interaction effect approached significance, Fs (1, 38) < 1. It is conceivable that imaginal encoding is the dominant means of processing graphic features for tagging purposes. If true, further instruction to use this strategy would be redundant.

An additional question concerns the distribution of errors across the full continuum of the distinctiveness variable. The overall correlation between distinctiveness ratings and errors per consonant approached statistical significance, ρ (20) = .42, $p < .10$. The failure to find a more pronounced correlation stems from the seemingly nonmonotonic covariation between distinctiveness and error rates. Grouping the consonants into successive sets of 4 consonants each, the total errors per set (ranging from most to least distinctive) were 42, 61, 43, 67, and 77. Thus, the departure from an otherwise regular increase in error rate with decreasing distinctiveness occurred for consonants of intermediate distinctiveness.

The confidence ratings revealed a pattern that paralleled that found for error rates, and they offered no additional insights into the processes underlying the effects of distinctiveness. Consequently, these data will not be reported here.

TABLE 1

Summary Statistics for Total Errors in Single-Item (Experiment I) and Multiple-Item (Experiment II) Recognition Learning

Distinctiveness	Instructions: Imagery M	Imagery SD	Standard M	Standard SD
Single-item recognition				
High	3.00	2.49	3.25	2.63
Low	4.40	2.85	3.85	2.28
Multiple-item recognition				
High	8.25	5.49	6.30	5.33
Low	13.15	8.66	9.95	8.16

Experiment II

Method

The design and general procedure followed that of Experiment I, but with appropriate modifications for the shift from a single-item to a multiple-item task. There were again 40 *S*s, 2 instructional conditions as a between-groups variable, and 2 levels of distinctiveness as a within-groups variable. Each experimental list contained 20 pairs of upper- and lowercase consonants (without additional filler pairs). One member of each pair was designated the right item. Half of the right items were uppercase forms (5 HD and 5 LD), the other half, lowercase forms (5 HD and 5 LD). Two variants of the list were employed, with the *right* and *wrong* members of pairs reversing functions across the variants. There were 8 study–test trials. On the study phase of each trial the individual pairs were exposed, via a Kodak Carousel projector, at a 2-sec. rate and .9-sec. interpair interval. The *right* member of each pair was underlined. Test pairs were presented in a test booklet, with a separate page for each trial. The *S*s indicated *right* items by encircling one member of each pair (self-paced). Different random orders were employed across study and test phases, and standard control procedures for verbal discrimination research were applied (e.g., *right* items appeared equally often on the spatial right and left for each study and test phase). Finally, to assure familiarity with the procedure, practice on the experimental list was preceded by 2 trials on a practice list composed of upper- and lowercase pairs of vowels.

Results and Discussion

Means and standard deviations for total errors on the 8 trials are given in Table 1. The main effect for distinctiveness was again statistically significant, F (1, 38) = 21.16, $p < .001$, with the error rate once more being greater for LD consonants than for HD consonants. As in Experiment I, neither the main effect for instructions nor the Instructions × Distinctiveness interaction effect approached significance, Fs (1, 38) = 1.60 and <1, respectively. If anything, there was a slight trend toward more errors under imagery instructions than under standard instructions.

The distribution of errors across the full continuum of the distinctiveness variable was examined further, this time in the context of multiple-item learning. The overall correlation between distinctiveness

ratings and errors per consonant attained statistical significance, ρ (20) = .50, $p < .05$. The total errors per set of 4 consonants (ranging from most to least distinctive) were 105, 157, 122, 170, and 199. Thus, the locus for the departure from a monotonic relationship coincided with that found for single-item recognition learning, i.e., with consonants of intermediate distinctiveness.

The extraction and concomitant tagging of distinctive graphemic features appear to be important components of both single-item and multiple-item recognition learning. It is this commonality of processes that presumably underlies the commonality of the functional relationships demonstrated in the present experiments. Thus, the present results add further support to the basic continuity between variants of recognition learning. In both variants, distinctive features of items are tagged, most likely by frequency-of-response cues (Kausler, 1973a, 1973b), during study phases. The tagged features are components of target items in the single-item task and components of *right* items in the multiple-item, or verbal discrimination, task. Subsequent discriminations are based on the presence of t[illegible] [illegible]ed features for target items, but n[illegible] [illegible]stractor items, in single-item tests, [illegible] [illegible]*ght* items, but not for *wrong* items, [illegible] [illegible]iple-item tests.

REF[illegible]

Anisfeld, M., & Kna[illegible], M. [illegible] [illegible]ssociation, synonymity, and dir[illegible] [illegible] false recognition. *Journal of Exper[illegible] Psychology*, 1968, **77**, 171-179.

Kausler, D. H. Hom[illegible] [illegible] double function items in verbal discrim[illegible] [illegible]arning. *Journal of Experimental Psy[illegible]*, [illegible]3, **98**, 214-216. (a)

Kausler, D. H. Ort[illegible] [illegible]inctiveness of homonyms and the [illegible] [illegible]ng hypothesis of verbal discrimin[illegible] *American Journal of Psychology*, [illegible] [illegible]-149. (b)

Kausler, D. H. Conti[illegible] [illegible] processes across variants of recogniti[illegible] [illegible]. In R. L. Solso (Ed.), *Theories [illegible] psychology: The Loyola Symposium*. [illegible], Md.: Lawrence Erlbaum, in press.

Raser, G. A. False re[illegible] [illegible] a function of encoding dimension an[illegible] [illegible]. *[illegible]ournal of Experimental Psychology*, 1972, [illegible]

Schulz, L. S., & Lovelace, E. A. Interpair acoustic and formal s[illegible] in verbal discrimination learning. *J[illegible] of Experimental Psychology*, 1972, **94**, 29[illegible]

Underwood, B. J. Fal[illegible] [illegible] produced by implicit verbal response[illegible] [illegible] *Experimental Psychology*, 1965, **70**, 122-1[illegible]

(Received May 16, 1973)

…nal of Experimental Psychology
74, Vol. 102, No. 3, 439–442

STUDY AND RETRIEVAL INTERVAL EFFECTS IN PAIRED-ASSOCIATE LEARNING

ALFRED A. BAUMEISTER[1] AND DORIS KISTLER

University of Alabama

Four groups of *S*s learned a list of paired associates (consonant–consonant–consonant: digit) under 4 combinations of study and retrieval times (2 vs. 6 sec.). Other *S*s were allowed to pace themselves during the study phase of the task. Transfer tests were carried out to measure cue selection and trigram integration. The major results were as follows: (*a*) study time up to 6 sec. influences rate of learning, (*b*) study time and retrieval time produce interactive effects on learning, (*c*) longer retrieval times are associated with a greater amount of single-letter selection and less stimulus integration, and (*d*) study time has little effect upon cue selection.

Studies of variations of item exposure duration have focused for the most part upon the rate rather than upon the process of acquisition. Much of the research pertaining to presentation interval has been generated by Bugelski's (1970) *total-time hypothesis* which states that the time required to learn a given amount of material is a constant regardless of how that time is spread across trials. The amount learned is a function of the total time committed to learning, within limits.

Actually, the total-time hypothesis is a generalized expression of a relationship between time available for learning and *performance*. It does little by way of elucidating what is learned and the way in which learning proceeds. Clearly, the manner in which time is distributed during a learning trial may differentially affect encoding, storage, and retrieval processes. Accordingly, the major focuses of the present investigation are upon the effects of the distribution of time in relation to cue selection and stimulus integration and the reciprocal interaction of acquisition and retrieval processes.

Much of the support for the total-time hypothesis has come from paired-associate (PA) learning studies, primarily of the anticipation variety. Cooper and Pantle (1967) noted that the functional time, insofar as the hypothesis is concerned, is probably the study interval. An implication of this conclusion is that rehearsal is the critical process affected by variations in study time.

Recently, a number of experiments have been reported that raise serious questions concerning the validity of the total-time hypothesis. Izawa (1971) has argued that the anticipation procedure of PA learning confounds the effects of study and test intervals. There is good reason to assume that acquisition and retrieval are qualitatively different processes. Izawa found that the way in which time is distributed in the study–test method produces effects that are different from those observed when anticipation procedures are employed. Stubin, Heimer, and Tatz (1970) also raised a number of conceptual and methodological criticisms of experiments that purportedly provide support for the total-time hypothesis. Using the study–test method, they reported that shorter presentation times are more efficient than longer presentation times. That is, doubling the number of exposures of a pair produced more effective performance than doubling exposure time per se. A similar conclusion was reached earlier by Johnson (1964).

The results of these and other investigations suggest the possibility that variations in study time may lead *S* to adopt different acquisition strategies. Furthermore, although acquisition and retrieval phases can be experimentally separated in the study–test method of PA learning, there is

[1] Requests for reprints should be sent to Alfred A. Baumeister, who is now at the John F. Kennedy Center for Research on Education and Human Development, George Peabody College for Teachers, Nashville, Tennessee 37203.

the possibility that time constraints placed on S during retrieval may affect his encoding of the stimuli.

In the present experiment, study time and retrieval time were varied factorially. The stimulus terms were consonant trigrams, the responses were digits. Transfer tests were made to determine degree of letter selection and trigram integration that occurred during learning.

Method

Subjects. The *S*s were 75 undergraduate volunteers from introductory psychology courses at the University of Alabama. None had been previously involved in PA learning experiments.

Materials. The learning materials consisted of 6 pairs with consonant trigrams as stimuli and single digits as responses. The pairs were as follows: PMD:7, KHB:5, VLT:6, GWS:4, JYC:8, and FRN:9.

Apparatus. The learning and transfer materials were presented to S on an inclined viewing console, 53.3 cm. long and 27.9 cm. high, in which 4 one-plane display projectors were mounted. The viewing surface for each projector was approximately 2.5 × 5 cm. Three projectors, containing letters, were mounted side by side. The fourth projector, containing the digits, was placed 5 cm. to the right of the other 3. A series of electronic timers and stepping relays were used to program presentation of the materials. All intervals and sequences were automatically controlled.

Procedure. The study–test (retrieval) method of PA learning was employed. Each of 60 *S*s was randomly assigned to 1 of 4 combinations of duration of study and retrieval intervals. The study and retrieval durations were either 2 or 6 sec., with the 4 combinations designated as 2–2, 2–6, 6–2, and 6–6. The number on the left refers to study interval, the number on the right to retrieval interval.

The remaining 15 *S*s were assigned to a self-paced study condition. The S held a switch that enabled him to change study items. These *S*s were given 2 sec. to retrieve the digit. Latencies were recorded for each study item.

Each S was read standard instructions for study–test PA learning. A complete trial consisted of one presentation of all 6 pairs and the 6 stimulus terms. A buzzer was used to cue S prior to the test phase within a trial. A constant 5-sec. interval separated study and test phases. Four different orders of items in both study and test were administered to prevent serial learning.

The criterion of learning was 2 successive perfect recitations of the list. Immediately upon reaching criterion, transfer tests were given. In the first transfer test, S was shown the 18 letters, 1 at a time, for 10 sec. each, and asked to recall the appropriate digit. Following digit recall, *S*s were again shown the 18 consonants and asked to identify the "missing" letters. Each was presented for 10 sec. Four different orders were used for individual letter presentation. The letters appeared on the console in the same position as in original learning.

Results

Original learning. Learning performance was expressed in terms of trials to criterion. Mean trials were 8.6, 6.1, 4.9, and 5.3 for the 2–2, 2–6, 6–2, and 6–6 groups, respectively. A 2 × 2 analysis of variance revealed a significant interaction effect, $F\ (1, 56) = 5.2$, $p < .05$. The *S*s learning under the short study and retrieval intervals (2–2) required more trials to reach criterion than *S*s in other groups. There was no significant difference between any of the other conditions.

The self-paced *S*s required a mean of 4.6 trials to attain the performance criterion. Analysis of variance involving just the groups given the 2-sec. retrieval interval revealed a significant effect, $F\ (2, 42) = 10.8$, $p < .01$. The 6-sec. and the self-paced groups did not differ and both were superior to the 2-sec. study group. The mean latency to respond during the study interval for the self-paced group was 15.2 sec/item. However, these *S*s were apparently able to differentiate learned from unlearned items in the study phase. They selectively took more time with items that they had failed on the previous trial,

TABLE 1

Mean Number of Single-Letter Solutions by Study Time and Retrieval Time

Retrieval	Letter position 1	2	3
	2-sec. study		
2 sec.	1.3	.2	.1
6 sec.	1.7	.2	0
	6-sec. study		
2 sec.	1.0	.1	.2
6 sec.	2.2	.1	.3
	Self-paced		
2 sec.	1.5	.1	.5

F (1, 14) = 10.2, $p < .05$. However, total time was not significantly correlated with trials to criterion (r = .08).

Cue selection. The double criterion method of scoring single-cue selection described by Berry and Baumeister (1971) was employed: A correct digit recall to one and only one of the letters of the trigram was scored as single-letter selection. Table 1 presents the mean number of single-letter solutions for various groups. A 2 × 2 × 3 analysis of variance including study time, retrieval time, and letter position was applied to these data (excluding the data from the self-paced condition). As expected, letter positions were differentially effective recall cues, F (2, 112) = 46.0, $p < .01$. More correct recalls were made to the first letter than to either of the other 2 letter positions. The main effect of retrieval time, F (1, 56) = 5.1, $p < .05$, and the Retrieval Time × Letter Position interaction, F (2, 112) = 3.6, $p < .05$, were significant. More single-first-letter solutions occurred under the 6-sec. retrieval interval than under the 2-sec. condition.

A separate 3 × 3 mixed analysis of variance comparing the self-paced, 2-sec., and 6-sec. study conditions across letter positions yielded a significant effect of position, F (2, 84) = 17.8, $p < .01$. Again, it was the first letter of the trigram that was the most effective recall cue.

Trigram integration. In the final phase of the experiment, *S*s were presented one letter of the trigram and requested to reproduce the remaining letters. The number of letters recalled was totaled for each individual. Means for the 2–2, 2–6, 6–2, and 6–6 groups were, respectively, 22.6, 17.1, 24.9, and 17.3 letters correctly recalled. An analysis of variance yielded a significant effect for retrieval time, F (1, 56) = 7.9, $p < .01$. This result can be interpreted to mean that significantly greater integration of the trigram occurred under the shorter retrieval interval.

The mean number of letters reproduced in the self-paced study condition was 21.5. An analysis of variance comparing the self-paced with the 2–2 and 6–2 groups failed to yield a significant effect, F (2, 42) = 1.49. Length of study time by itself seems to have little effect on trigram integration.

Discussion

The results demonstrate that variations of study time and retrieval time may produce interactive effects upon PA performance. The *S*s given 6 sec. to study each pair reached criterion faster than those who were given 2 sec. of study time. However, in the case of those individuals who viewed the pair for only 2 sec. but who also were provided relatively long (6 vs. 2 sec.) retrieval time, the detrimental effects of the brief study interval were mitigated considerably. This finding implies the existence of some sort of information feedback mechanism from retrieval to storage process in PA learning. As time constraints are relaxed in the retrieval phase, different strategies appear to be utilized during storage. Loftus and Wickens (1970) have reached a similar conclusion based upon their study of incentive cuing in the retrieval phase of PA learning.

Tests for letter selection and letter reproduction indicate that variations in retrieval interval influence encoding strategies. In particular, it appears that, given greater time to retrieve the response term, college students are more inclined to select the first letter of a consonant trigram as the functional cue. Not only was there a greater amount of single-letter selection in the 6-sec. retrieval condition, relative to the 2-sec. condition, but there was also less integration of the stimulus compound under the former condition.

Although these results may, at first, seem to be at variance with what one might expect on a commonsense basis, we believe that when the test or retrieval phase is brief, *S*s do not have sufficient time to actively adopt a cue-selection strategy. What is learned during the acquisition phase in the PA task is influenced by time constraints placed on recall.

The performance of the self-paced *S*s is instructive in this regard. Although they were allowed all the time they wished to study each pair, they were permitted only 2 sec. for retrieval of the response. Transfer tests for digit recall to single letters and for letter reproduction revealed that the self-paced group performed in much the same ways as the other E-paced *S*s in the 2-sec. test condition. Lovelace and Greenberg (1969) have likewise reported that the length of the study interval

(2 vs. 6 sec.) does not affect cue selection in PA learning with consonantal trigrams as stimuli.

Any hypothesis stating that the amount learned is a function of the time given to study is an oversimplification on at least 2 counts. First, study time and retrieval processes are not independent. Secondly, even if overall performance is the same when time is equated, different learning strategies may be employed.

REFERENCES

Berry, F. M., & Baumeister, A. A. Measuring single-letter cue selection in paired-associate learning: A methodological note. *Journal of Psychology*, 1971, **77**, 181–187.

Bugelski, B. R. Presentation time and the total-time hypothesis: A methodological amendment. *Journal of Experimental Psychology*, 1970, **84**, 529–530.

Cooper, E. H., & Pantle, A. J. The total-time hypothesis in verbal learning. *Psychological Bulletin*, 1967, **68**, 221–234.

Izawa, C. Massed and spaced practice in paired-associate learning: List versus item distributions. *Journal of Experimental Psychology*, 1971, **89**, 10–21.

Johnson, N. F. The functional relationship between amount learned frequency vs. rate vs. total time of exposure of verbal materials. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 502–504.

Loftus, G. R., & Wickens, T. D. Effect of incentive on storage and retrieval processes. *Journal of Experimental Psychology*, 1970, 85, 141–147.

Lovelace, E. A., & Greenberg, B. Cue selection in paired-associate learning: Length of study interval. *American Journal of Psychology*, 1969, **82**, 254–258.

Stubin, E. J., Heimer, W. I., & Katz, S. J. Total time and presentation time in paired-associate learning. *Journal of Experimental Psychology*, 1970, **84**, 308–310.

(Received May 18 [illegible]

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 443–446

ATTENTION BIAS AND THE RELATION OF PERCEPTION LAG TO SIMPLE REACTION TIME [1]

A. J. SANFORD [2]

University of Dundee, Dundee, Scotland

Previous research suggests that stimulus loudness affects simple reaction time to a greater extent than it affects subjective simultaneity judgments. A question has recently arisen as to whether these results can be attributed to an attention bias like that of "prior entry." An experiment precluding the possibility of this artifact replicated previous results. Some alternative interpretations of the results are briefly outlined.

There is evidence to suggest that a simple phenomenological model is adequate to account for the relationship between visual intensity and simple reaction time (RT). This view supposes that perceptual lag, i.e., the time to perceive a stimulus from its onset, is influenced by stimulus brightness, and that a reaction is initiated as soon as perception of the stimulus has occurred (Roufs, 1963, 1966). That is to say, the effect of stimulus intensity changes on difference in the time of perception is the same as the effect of RT differences. In studies supporting this contention, Roufs (1963) compared the size of the effect of intensity on behavioral RT and phenomenal simultaneity and found that the size of the intensity effect was the same in both cases. Others have suggested that not all of the behavioral RT effect is accounted for by perception lag (Hohle, 1967). This is particularly evident in the case of auditory intensity (loudness). There is evidence that RT is affected to a greater extent by loudness changes than is perceptual latency, as revealed by a perceptual simultaneity task (Sanford, 1971). Indeed, Roufs (1963) has reported a failure to find any effect of auditory intensity on the point of subjective simultaneity for click–flash pairs.

The purpose of the present article is to examine the apparent discrepancy between the effect of loudness on perceptual lag and simple RT using the same paradigm as that of Sanford (1971). Specifically, the design used is an attempt to ensure that the conclusions that have been drawn are not based upon data resulting from an attentional artifact. The task used in the previous study required *S* to watch a quickly rotating pointer move across a circular dial. While *S* was watching the display, an auditory signal was presented, and *S*'s task was to say where the pointer was positioned at the onset of the sound. By manipulating the intensity of the sound, some changes in the reported position of the pointer at onset were obtained, and the results suggested that perceptual lag increased with decreases in stimulus intensity. The increase was, however, much smaller than the increase in behavioral RT measured for the same loudness range. It was argued that, with such a design, behavioral RT and perception lag could be measured under almost identical conditions, thereby minimizing the possibility of differences in intensity effect magnitude resulting from different response criteria. (See Murray, 1970, or Grice, 1968, for a discussion of the response criterion concept.)

The design actually employed entailed the presentation of the different intensities in blocks. However, this may have had a biasing effect on the distribution of *S*'s attention to the two components (visual and auditory) of the task. In particular, it may be argued that the trial blocks in

[1] This research was undertaken in partial fulfillment of the PhD degree requirements at the University of Cambridge while the author held a Medical Research Council studentship at the Medical Research Council Applied Psychology Unit.

[2] Requests for reprints should be sent to A. J. Sanford, Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland.

TABLE 1

RELATION OF MEAN REACTION TIME AND MEAN L VALUES (IN MSEC.) TO STIMULUS-INTENSITY INCREASE

Measure	Stimulus (db. above masker)				Range (maximum)
	2	3	7	18	
Reaction time					
M	314	231	198	184	130
SD	25	28	35	30	
L					
M	61	37	29	31	32
SD	21	13	15	15	

Note. The L value is derived from the actual pointer position − the observed pointer position.

which extremely close-to-threshold stimuli were employed would cause S to attend more closely to the auditory display than he would for higher signal-to-noise ratios. Although we are only just beginning to construct models of attention bias of this type (cf. Sternberg & Knoll, 1973), the idea that attention bias influences the perceptual lag has long been in existence as the law of prior entry (Titchener, 1908), which states: "The stimulus for which we are predisposed requires less time than a like stimulus, for which we are unprepared to produce its full conscious effect [p. 251]." Stone (1926) and Sternberg and Knoll have demonstrated that simultaneity judgments are influenced by attentional bias in the manner predicted by the prior-entry law.

In the case of Sanford's (1971) pointer test it may be argued that the prior-entry effect could produce the result that the intensity effect under the pointer test was less than that under RT conditions. When blocks of quiet stimuli are being presented, additional attention may be paid to the auditory stimulus, which would, by prior entry, cause the stimulus to be detected earlier than it would were less concentration being devoted to it. Thus, attentional bias runs against the naturally greater detection lag that could be associated with quiet stimuli. When blocks of loud stimuli are used, the converse argument applies. Although detection lags are smaller for loud stimuli, this effect may be offset by less attention being paid to the loud stimuli. The result of these two biases would be to lessen the magnitude of the intensity effect in the clock task. Such a possible bias can only be eliminated by using a design in which the stimuli of differing intensities are presented randomly.

METHOD

Apparatus. The S ($N = 8$ Royal Navy enlisted men) was required to do two tasks, the RT task and pointer task. In the pointer task, Ss had to watch a pointer revolving around a clock face and say just where the pointer was at the time when the onset of the stimulus occurred. The "clock" was a chronoscope graduated in centiseconds, with a 1-sec. sweep. Every tenth-of-a-second was numbered 0, 10, 20 . . . 90. The sweep-face was 8-cm. in diameter, and the face was set centrally into a 14.5-cm.^{2} box. The whole chronoscope was set into a 50-cm.^{2} board. The board was mounted on a table in front of S at a viewing distance of 60 cm. A comfortably bright level of illumination was used and lamps were arranged so that variations over the surface of the clock face were minimal. Stimuli consisted of increments in a background of random noise—intensities, measured by a Dawe sound-level meter coupled to the earphones, were 60 db. for the background rising to 62, 63, 67, and 78 db. for Stimuli 1, 2, 3, and 4. Stimuli were effectively continuous, being terminated by E after S had responded. They were presented binaurally, and had an effectively zero rise time.

On each trial, S watched the pointer, which swept over the clock face once, and a stimulus was then introduced on the second sweep. Temporal uncertainty was thus in the range 1–2 sec. The distribution of stimulus onset points was roughly equally dispersed across the range 1–2 sec.

The RT apparatus consisted of a microswitch arranged on the release-to-react principle for S's nonpreferred hand. Stimulus onset activated a system of decatron timers which ceased counting when S reacted.

Procedure. There were three sessions. In the first session, Ss were acquainted with the stimuli and given some practice at the RT task. In the following two sessions, Ss heard 30 stimuli at each intensity (120 stimuli in all) in a random order. In addition, 24 blank trials were introduced in which no stimulus was presented. The first of these sessions was considered practice. All analyses are therefore based on final-session data. The Ss were instructed to perform the pointer test and the RT test simultaneously, and were told that the pointer test was concerned with visual acuity, and that they were to say where the pointer was when the stimulus first came on. Instructions for RT were standard.

RESULTS

The pointer test scores were made in the following way. A value of *L* (actual pointer position minus the observed pointer position), was obtained for each judgment. The mean of these values was then obtained. The RT and *L* scores are given in Figure 1 and Table 1 and demonstrate the attenuated intensity function obtained in the judgment task.

A simple nonparametric test, Friedman's analysis of variance (Siegel, 1956), was applied to each set of scores and revealed an effect of intensity on RT ($p < .001$) and an effect of intensity on *L* ($p < .001$). The interaction effect was tested by comparing RT–*L* values as a function of intensity and was significant

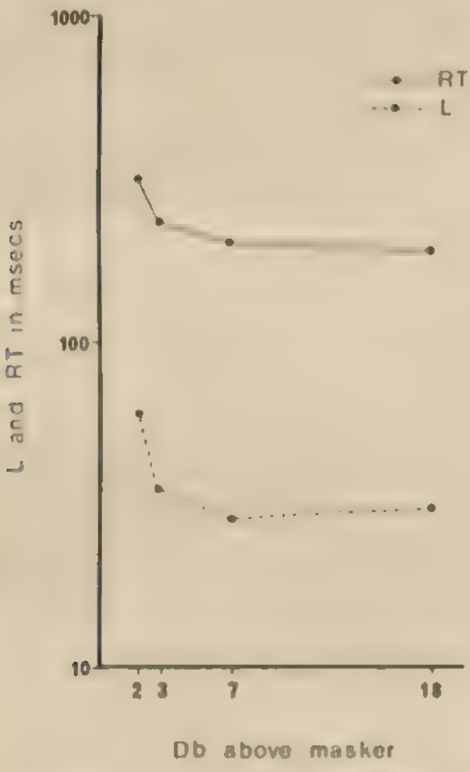


FIGURE 1. Mean values of reaction time (RT) and *L* (actual pointer position − observed pointer position) plotted on a log scale against stimulus intensity. (It is apparent from this that the change in RT is greater than the change in *L*. Statistical analysis was carried out on the untransformed data.)

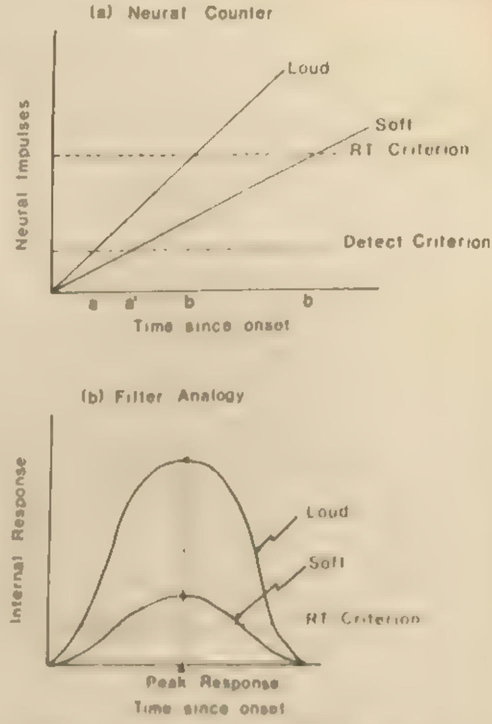


FIGURE 2. (a) Graphical representation of the counter model (e.g., Grice, 1968). (The effect of intensity on the detect-for-judgment [DJ] state, a′ a, is seen to be less than that on the detect-to-react [DR] state, b′ b, because the DJ state criterion requires less evidence than the DR state criterion.) (b) The internal response to a stimulus can also be considered as analogous to the output of a simple filter (Sternberg & Knoll, 1973). (If an event such as the peak internal response is used to trigger DJ states, then if there is any effect of intensity at all it should be less than that obtained in reaction time [RT] tasks, since DR states would be initiated by a variance setting amplitude detector analogous to the criterion of the counter model.)

($p < .001$). Hence, the results show that although intensity influenced *L* and RT, RT was influenced to a much greater extent. (Parametric tests were not used since use of a transform to satisfy parametric assumptions would be incompatible with considering a specified scale of measurement, namely, linear time. The primary interest in this experiment is in the interaction of RT and *L*, and this would be modified by any nonlinear transformation.)

Discussion

The results obtained in the present experiment are substantially similar to those obtained by Sanford (1971). They support the view that the differential effects of stimulus intensity on RT and simultaneity cannot be attributed to an attentional bias interacting with the intensity effect, since this was prevented by the use of random stimulus schedules.

The most general statement which can be made on the basis of the results is that different amounts of stimulus information have to be sampled in order to cause a detect-to-react (DR) and detect-for-judgment (DJ) event to occur. Sanford (1972) suggested that the DR state may well occur at an earlier date than the DJ state. This is based on the argument used in favor of the simple counter model of RT, i.e., that smaller intensity effects reflect less stimulus sampling than larger ones. This is consistent with the view of Murray (1970) and others that the magnitude of an intensity effect is directly related to the lag in stimulus detection, since it reflects criterion level. This is clear from the model shown in Figure 2a where neural impulses are assumed to be counted until a criterion number have been accumulated. Quiet signals are assumed to cause slower neural impulse rates than loud ones. This interpretation is consistent with the findings that *S*s are aware of some of the intensity-dependent delay in RT, when measures are made of subjective ratings of RTs (Sanford, 1970).

Other experimental work comparing intensity functions for RT and simultaneity has recently been discussed by Sternberg and Knoll (1973), who have developed an alternative suggestion that DJ states may be triggered at a later point on the stimulus continuum than DR states. They suggest that the internal representation of a stimulus may be likened to the response of a filter to a step input, Figure 2b. The occurrence of a particular level of input mediates the DRs, while DJs are based on the peak response of the system. Since peak responses occur at the same time for stimuli of different amplitudes, an absence of intensity effect on tasks tapping the DJ system might be expected. However, Sternberg and Knoll do suggest this model for pulsed stimuli only, presumably with a duration less than the time constant of the filter. Whether it can be generalized to stimuli longer than the time constant (as used in the present study) is an open question. If it could be so generalized, it would reflect a detection of asymptote rather than peak. It is, of course, possible that the asymptote of the filter output may be associated with some neural event which could act as a cue for DJ states if the filter model is correct. At this stage, it seems impossible to distinguish between these two views, and Sternberg and Knoll point out that DR and DJ states may be elicited by different aspects of the internal representation of the stimulus in tasks that differ only slightly. The present results merely show that the discrepancy between RT and simultaneity measures of detection latency cannot be accounted for by an explanation based upon attentional bias.

References

Grice, G. R. Stimulus intensity and response evocation. *Psychological Review*, 1968, **75**, 359–373.

Hohle, R. H. Component process latencies in reaction times of children and adults. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior.* Vol. 3. New York: Academic Press, 1967.

Murray, H. G. Stimulus intensity and reaction times: Evaluation of a decision-theory model. *Journal of Experimental Psychology*, 1970, **84**, 383–391.

Roufs, J. A. J. Perception lag as a function of stimulus luminance. *Vision Research*, 1963, **3**, 81–91.

Roufs, J. A. J. On the relation between the threshold of short flashes, the flicker fusion frequency and the visual latency. In A. van Katwijk (Ed.), *IPO Annual Progress Report.* 1. Eindhoven, The Netherlands: Institute for Perception Research, 1966.

Sanford, A. J. Rating the speed of a simple reaction. *Psychonomic Science*, 1970, **21**, 333–334.

Sanford, A. J. Effects of changes in the intensity of white noise on simultaneity judgements and simple reaction time. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 296–303.

Sanford, A. J. Loudness and simple reaction time. *Sound*, 1972, **6**, 92–95.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Sternberg, S., & Knoll, R. L. The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and Performance.* IV. New York: Academic Press, 1973.

Stone, S. A. Prior entry and the auditory-tactual complication. *American Journal of Psychology*, 1926, **37**, 284–287.

Titchener, E. B. *Lectures on the elementary psychology of feeling and attention.* New York: Macmillan, 1908.

(Received May 21, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 447–455

ASYMMETRIES IN JUDGMENTS OF VERTICALITY[1]

PHILIP H. K. SEYMOUR[2]
University of Dundee, Dundee, Scotland

Two experiments are reported in which *S*s judged whether or not the word *above* or *below* correctly specified the location of a dot relative to a picture of a face. In Experiment I, the face was horizontally oriented with its top to the left or right and the dot to the left or right. An analysis of the latency data showed a substantial *same–different* effect and a marginally significant advantage for *above* vs. *below*. In Experiment II, the face was normally oriented or inverted. The normal face was associated with a significant *above* vs. *below* difference and a *same–different* effect. When the face was inverted, both of these effects were eliminated. These results are discussed in relation to recent theories about the verification of spatial locations.

Seymour (1969) reported a study in which *S*s made *same–different* judgments of displays consisting of the words *above* or *below* and pictures of a dot above or below a reference square. The main findings were that reaction times (RTs) were faster for *above* displays than for *below* displays, and faster for *yes* reports than for *no* reports. These findings were subsequently confirmed by Chase and Clark (1971).

Chase and Clark (1971) argued that the *above–below* asymmetry must depend on differences in encoding time for the words *above* and *below* and not on a tendency to scan the display in a top–down direction as had been proposed by Seymour (1969). However, the evidence cited by Chase and Clark was not sufficient to determine this issue. In their Experiment 2 they showed that the *above–below* asymmetry disappeared when the words were replaced by arrows; this alters the task to one of 2-choice responding to the proximity of a dot and arrow/head and appears not to involve a proper comparison of locations. In Experiment 3 they replaced the words *above* and *below* by *present* and *absent*, which were to be matched against a dot that might or might not appear at a constant location within a block of trials; this experiment would bear on the question of focused attention on the upper region of the figure only if the dot signifying present were allowed to occur randomly in the upper or lower locations.

Chase and Clark's (1971) Experiment 4 involved figures in which the dot occurred only in the above location for one group and only in the below location for the other. The *S*s were instructed to infer the below location from *absent*–above in the one case, and above from *absent*–below in the other. In the top-visible condition, the pattern of latencies was similar to that obtained under the normal condition, but substantial alterations occurred in the bottom-visible condition, which Chase and Clark interpreted as an addition to the time required to encode the location above. On the other hand, the top-visible condition did not result in an increase in the time to encode the location below, so that Chase and Clark commented that, "Apparently, subjects normally attend to the upper location of the circle, hence they normally (or often) only *infer* (*circle below square*) in the *BELOW* displays [p. 322]."

Given this latter comment, it appears that Seymour (1969) and Chase and Clark (1971) are in agreement in arguing that attention is often initially focused on the upper region of the figure; a tendency of this kind might delay encoding of *below* displays because of top–down scanning or because of the occurrence of a time-consuming inference. Chase and Clark have added the assumption of a difference in encoding time for the words *above* and *below*, and have argued that this occurs because

[1] This research was supported by Grant HR/1787/1, from the Social Science Research Council of the United Kingdom. The author is grateful to Jenny Greenhalgh for her assistance in obtaining and analyzing the data.

[2] Requests for reprints should be sent to Philip H. K. Seymour, Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland.

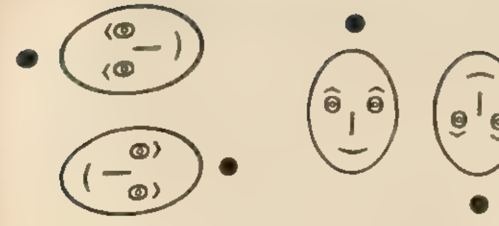


FIGURE 1. Examples of face–dot displays used for (a) horizontal conditions of Experiment I and (b) vertical conditions of Experiment II. (All 4 pictures require the response *yes* when presented with the word *above*, and *no* when presented with the word *below*.)

the word *below* refers to a negative pole on a dimension of verticality (Clark & Chase, 1972). Thus, the word *above* may be represented as + vert/ + polar and the word *below* as + vert/ − polar, indicating that these words are semantically represented as sharing values on a verticality feature, but as referring to positive and negative poles of the vertical dimension (Leech, 1969). Chase and Clark (1971) argued that the −polar case involves some addition to the sentence-encoding stage of the comparison.

The introduction of the notion that *below* is an inherently negative concept suggests a further line of theorizing which may be relevant to interpretation of the *above–below* asymmetry. Schaeffer and Wallace (1970) described a model for comparisons of word meanings in which semantic similarity, as determined by a preliminary holistic comparison of 2 word concepts, affects the thresholds of *same* and *different* counters used to test for equivalence of the 2 concepts with respect to some specified attribute. Semantic similarity tends to lower thresholds for *same* judgments and to raise thresholds for *different* judgments; conversely, semantic dissimilarity raises the threshold for *same* judgments, but lowers the threshold for *different* judgments. Seymour (1973) proposed that the inherent positivity or negativity of the concepts to be compared might have effects analogous to those of semantic similarity. Thus, if the representation of an *above* above word–picture display is inherently positive, as Clark argued, this display may result in a greater reduction in the threshold for a *yes* response than a *below*–below display. As a preliminary test of this assumption, Seymour instructed *S*s to report *no* to *above*–above and *below*–below displays, and *yes* for *above*–below and *below*–above displays. This reversal of the congruence-report mapping eliminated both the *above* below and the *true–false* effects. It was argued that this result was incompatible with the assumptions of the Chase and Clark (1971) model, but might arise if the advantage of *above*-above displays depended on the possibility of mapping the display onto a semantically affirmative report.

It seems possible, therefore, to identify 3 general theories about the nature of the asymmetry observed in verticality judgments. (*a*) A *perceptual* theory states that *S* scans the picture display from top to bottom, or attends initially to the upper region of the figure, so that encoding time is less for above pictures than for below pictures. This is the theory which was attacked by Chase and Clark (1971), although, as mentioned above, they restated an attentional variant of it in the discussion of their Experiment 4. (*b*) A *linguistic* theory states that the concepts *above* and *below* differ in semantic complexity, and that the marked term, *below*, requires more time at the word-encoding stage than the unmarked term, *above*. Pictures are encoded in the same abstract or propositional format as sentences, so that a comparable difference in encoding time is postulated for above vs. below pictures. (*c*) A *response availability* theory states that positive features in the semantic representation of the word–picture display tend to reduce the threshold for selection of a *yes* response. Responses to *above*–above displays are faster than responses to *below*–below displays because there is greater inherent positivity in the representation of *above*–above.

This article describes 2 experiments which were designed to discriminate among these 3 theories.

EXPERIMENT I

Experiment I aimed to distinguish between the perceptual and linguistic theories.

The *Ss* matched the words *above* or *below* against the location of a dot which was positioned relative to a horizontally oriented reference face. Figure 1(a) shows the face–dot configurations in which the dot was located above the face. In the below displays, the dot was located close to the chin of the face. Thus, the above and below locations might be to right or left of the display, depending on the position of the top of the face. This type of display eliminates any possible advantage for focusing attention on the upper part of the figure, and controls for tendencies to focus attention to the left or right. If the *above–below* asymmetry disappears under these conditions, it will follow that the effect is largely attributable to a distinction between the above and below locations, or spatial/perceptual factors. If the asymmetry persists, this will count as support for Clark's contention that *above* concepts may be established more rapidly than *below* concepts, but will also be compatible with the assumptions of the response availability theory.

Method

Subjects. Eight volunteers from final year classes at Dundee, Scotland, senior secondary schools served as unpaid *Ss*.

Apparatus. On each trial a warning tone was sounded for 500 msec. followed by a 1-sec. delay after which the word and picture displays were rear-projected onto a screen for 2 sec. The light sources of 2 projectors were obscured by a leaf shutter, which could be deflected by pulsing a galvanometer mounted on top of the projector. A Venner clock started when the projector shutters were opened, and stopped when *S*'s vocal *yes* or *no* response closed the relay of a voice key.

Displays. The words *above* and *below* were printed in lowercase Letraset (Univers 55) on cards and then photographed and mounted as slides. The pictures were faces similar to those shown in Figure 1, with a dot above the head or below the chin. The ellipse outline of the face was 5 cm. in length, and the center of the dot was 1 cm. from the circumference of the ellipse. These pictures were also photographed, and 2 slides were made for each picture, so that the face could be presented in horizontal orientation, with the top of the head to the left or right of *S*, and the dot located at the top of the head or at the chin. On each trial, *S* saw the word *above* or *below* displayed above 1 of the 4 face pictures. Under the conditions of the experiment, the visual angles subtended by the word and picture were 2° and 3°, respectively.

TABLE 1

MEAN RESPONSE LATENCIES (IN MSEC.) OF *yes* AND *no* REACTIONS TO *above* AND *below* DISPLAYS WHEN THE REFERENCE FACE WAS HORIZONTALLY ORIENTED: EXPERIMENT I

Head direction	*above*–above	*below*–below	*above*–below	*below*–above
Right				
M	976	1,055	1,122	1,114
SE_M	55	70	74	55
Left				
M	985	1,039	1,081	1,134
SE_M	63	77	78	91
M	981	1,047	1,102	1,124

Procedure. All *Ss* started with some practice in the use of the voice key and in classification of the displays. They were instructed that they should respond *yes* as rapidly as possible when the word display correctly specified the position of the dot relative to the face, and *no* when it did not. The main test session was organized as 3 blocks of 32 observations. Each of 8 word–picture combinations occurred 4 times per block in a random order independently determined for each *S* at each block. Error responses were discarded and replaced. Each 30–45-min. testing session was carried out in a quiet darkened room.

Results and Discussion

Error rates in this experiment were quite low, amounting to only 2.5% of all trials, and were not related in any consistent way to the main conditions of the experiment.

A summary of the latency data is given in Table 1, classified by word–picture combination (*above*–above, *below*–below, *above*–below, and *below*–above) and location of the dot (left or right). Each mean is based on 12 observations on each of 8 *Ss*. A preliminary analysis of variance established that the effects for dot locations were not significant, either for *yes* or for *no* responses, *F*s $(1, 7) < 1$ and 1.6, and that the interaction of this factor with effects for the words, *above* or *below*, also was not significant, $F\ (1, 7) < 1$.

The RTs were collapsed across dot position, and an analysis of the type employed by Chase and Clark (1971) was run to test the effects of response (*yes* vs. *no*) and word (*above* vs. *below*). This analysis indicated that *same* combinations were verified about 100 msec. faster than *different* combinations, $F\ (1, 7) = 59.89$, $p < .001$. Neither the *above–below* differ-

TABLE 2

Possible Semantic Representation of *above below* Displays When the Reference Face Was Horizontally Oriented

Item	*above*–above	*below*–below	*above*–below	*below*–above
Word	+vert/+polar	+vert/−polar	+vert/+polar	+vert/−polar
Dot	−vert/→right	−vert/←left	−vert/←left	−vert/→right
Face	−vert/→right	−vert/→right	−vert/→right	−vert/→right
Dot/face top	+match	−match	−match	[illegible]
Word/Dot/face–top	+match	+match	−match	−match
Decision	+true	+true	−true	−true
Summed positivity	5	3	2	2
Summed negativity	2	4	5	5
Neutral components	2	2	2	2

ence of 44 msec., F (1, 7) = 4.14, nor the Response × Word interaction, F (1, 7) = 1.67, was significant. In a similar analysis, in which response and above vs. below pictures were factors, neither the pictures effect, F (1, 7) = 1.67, nor the Response × Pictures interaction, F (1, 7) = 4.14, was significant.

These results contrast with those obtained for vertically oriented displays by Seymour (1969) and Chase and Clark (1971). Both previous studies showed highly significant effects for *above* vs. *below* and for *same* vs. *different*. In the present experiment, there was a large *same–different* effect, but the *above–below* effect was not significant. Hence, if one adopts the assumptions about independence and additivity of component encoding and comparison stages stated by Chase and Clark (1971, 1972) and Clark and Chase (1972), an appropriate conclusion might be that the *above–below* effect is dependent on the vertical arrangement of the picture display, but that the *same–different* effect is independent of this factor. This conclusion is contrary to the predictions of the linguistic theory, which states that encoding time is shorter for *above* sentences and above locations than for *below* sentences and below locations. Use of a horizontal reference face might be expected to add a small constant to the duration of the picture-encoding stage. It should not, however, affect the difference in encoding time as between *above* and *below* sentences, or above and below pictures, which, in the theory, are supposed to reflect the time required to set up abstract propositional representations.

The main prediction of the perceptual theory appears to be that there should be no difference between the *same* displays, *above*–above and *below*–below, or between the *different* displays, *above*–below and *below*–above, although there may be an overall *same–different* effect. This would follow if the *above–below* effect depended on location of the critical dot in the upper region of the figure. In the case of *no* responses, the differences between *above* and *below* displays were not significant, F (1, 7) < 1. However, in the case of *yes* responses, the difference of 66 msec. between *above*–above and *below*–below was significant, F (1, 7) = 6.46, p <.05. The outcome for *no* responses is consistent with the perceptual theory, therefore, but the outcome for *yes* responses is not.

In order to consider the predictions of the response availability model, it is necessary to make a number of assumptions about the nature of the semantic representation of the word–picture display. A possible formulation has been outlined in Table 2. It has been assumed that the encoding of the display involves establishment of featural descriptions of (a) the word *above* or *below*, (b) the absolute location of the dot (left or right), (c) the absolute location of the top of the face (left or right), (d) the outcome of a comparison of dot location and face–top location, (e) a comparison of word polarity and match or mismatch of the dot and face–top locations, and (f) an indication of truth or falsity of the display. The form of the featural description for word meaning, and dot and face–top location, is based on semantic analyses of locatives proposed by Leech (1969). The

TABLE 3

Possible Semantic Representation of Normal and Inverted Face *above–below* Displays

Item	Normal				Inverted			
	above–above	*below*–below	*above*–below	*below*–above	*above*–above	*below*–below	*above*–below	*below*–above
Word	+vert/ +polar	+vert/ −polar	+vert/ +polar	+vert/ −polar	+vert/ +polar	+vert/ −polar	+vert/ +polar	+vert/ −polar
Dot	+vert/ +polar	+vert/ −polar	+vert/ −polar	+vert/ +polar	+vert/ −polar	+vert/ +polar	+vert/ +polar	+vert/ −polar
Face	+vert/ +polar	+vert/ +polar	+vert/ +polar	+vert/ +polar	+vert/ −polar	+vert/ −polar	+vert/ −polar	+vert/ −polar
Dot/face-top	+match	−match	−match	+match	+match	−match	−match	+match
Word/Dot, face-top	+match	+match	−match	−match	+match	+match	−match	−match
Decision	+true	+true	−true	−true	+true	+true	−true	−true
Summed positivity	9	6	5	6	7	6	5	4
Summed negativity	0	3	4	3	2	3	4	5

vertical dimension is characterized by the feature +vert, and the horizontal dimension by the feature −vert. The upper region of the vertical dimension is defined as positive and has the feature +polar. The lower region is negative and is assigned the feature −polar. Since the horizontal dimension is probably not bipolar in this sense, the directions left and right have been represented by the neutral features ← left and → right. The outcomes of the various comparison operations are represented by the features +match and −match.

It is assumed that as *S* encodes the display a featural representation of the type suggested in each column of Table 2 is constructed. The response to be made depends ultimately on inspection of the value for the +match or −match feature, which defines the outcome of the comparison of word polarity and dot/face–top congruity. This corresponds to the test for sameness or difference of value in relation to a specified attribute in the model of Schaeffer and Wallace (1970). It is further assumed that, as the semantic representation of the display is constructed, positive feature values tend to reduce the threshold for initiation of a *yes* response, whereas negative feature values will reduce the threshold for a *no* response.

The last 3 rows of Table 2 show the values obtained by making a simple count of positive, negative, and neutral features for the 4 displays. Summed positivity for an *above*–above display has a value of 5 vs. a value of 3 for a *below*–below display Hence, the reduction in threshold for the *yes* response is expected to be somewhat greater for *above*–above than for *below*–below, and this might give rise to the small but significant difference in RT that was observed. In the case of *no* responses, summed negativity has a value of 5 for both *above*–below and *below*–above displays. Thus, these 2 displays will be expected to reduce the threshold for a *no* response by an equivalent amount, and no difference in RT is expected.

Experiment II

Experiment II was carried out with the aim of determining the effects on the *above–below* and *true–false* differences of inversion of the reference face. The words *above* and *below* were matched against the location of a dot relative to a reference face which might be normally oriented, or inverted, as shown in Figure 1(b). The normal orientation of the face corresponds to the condition tested in the studies by Seymour (1969) and Chase and Clark (1971), and may be expected to reproduce the significant *above–below* and *same–different* effects obtained in those studies. However, the 3 theoretical positions discussed earlier lead to different predictions for the case where the reference face is inverted.

If, as is implied by the perceptual theory, the *above–below* asymmetry depends on a tendency to focus attention on the top of the figure or to scan the figure in a top–down direction, inversion of the face should

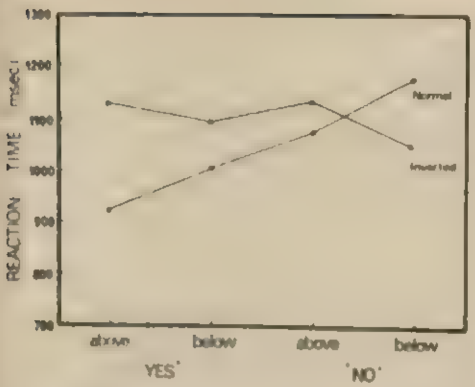


FIGURE 2. Summary of mean reaction times (in msec.) for classification of vertical face displays (Experiment II). (The means are for *yes* and *no* responses to *above* and *below* displays when the face might be normally oriented or inverted.)

result in a symmetric reversal of the *above–below* effect. This assumes that *S* may take some time to note the inversion of the face, but that, once this has occurred, he will encounter the evidence critical for verification of the display sooner for a *below*–below combination than for an *above*–above combination. Since these effects are arising at the stage of picture encoding, the *same–different* effect should not be altered.

An assumption of the linguistic theory is that an abstract representation of "X above Y" may be set up more rapidly than a representation of "X below Y," for both sentences and pictures. If sentence encoding and picture encoding are genuinely independent mental operations, inversion of the reference face will not alter the difference in encoding time as between the words *above* and *below*. The effect of inversion should be on the duration of the picture-encoding stage, and should also be a constant increment which will leave the difference in encoding time for *above* vs. *below* pictures unaltered. Further, if the encoding stages are independent of the comparison stage, inversion of the face should not influence the magnitude of the *same–different* effect. In general, therefore, the linguistic theory predicts a constant increment in RT when the face is inverted, but no change in the basic *above–below* and *true–false* effects.

The predictions of the response availability model must again depend on an assessment of the balance of positive and negative components in the semantic representation of the displays. Table 3 gives the outcome of an analysis which is similar in form to that proposed for the horizontal faces. Word meaning, dot location, and face orientation are represented by a +vert feature and a +polar or −polar feature, and the outcomes of comparisons of dot location and face–top location, or word polarity and congruence of dot and face–top, by +match or −match features. The last 2 rows of the table show the values for summed positivity and summed negativity obtained for each display. For normal orientation of the face, there are 9 positive components in the representation of an *above*–above display, but only 6 in the representation of a *below*–below display. However, when the face is inverted, the ratio of positive components becomes 7:6. The prediction of the response availability model for *yes* responses, therefore, is that there will be a significant difference between *above*–above and *below*–below displays when the face is normal, but that this difference will be much attenuated when the face is inverted. In the case of *no* responses, the ratio of negative components is 4:3 in favor of *above*–below for normal orientation of the fact, but 5:4 in favor of *below*–above when the face is inverted. The best prediction, therefore, is that *above*–below will be classified faster than *below*–above when the face is normal, but that the direction of this difference will be reversed when the face is inverted.

Method

Subjects. Eight volunteers, all University of Dundee students, served as unpaid *Ss*.

Apparatus and displays. These were identical to Experiment I except that the faces were now displayed in a normal or inverted orientation, and the word display appeared to the left of the picture on each trial.

Procedure. The procedure was similar to that followed in Experiment I. After some preliminary practice, each *S* recorded a total of 96 observations in which each of 8 word–picture combinations occurred 12 times in random orders independently determined for each *S*. Error responses were discarded and replaced.

Results and Discussion

Error rates were again low in this experiment, being limited to 2.8% of all trials.

Figure 2 gives a summary of the latency data for the 4 word–picture combinations and for normal or inverted orientation of the reference face. A preliminary analysis was carried out to assess the effects of inversion of the face in relation to differences among the word–picture combinations. This analysis indicated that inverted face displays were classified about 58 msec. slower than normal face displays, $F\ (1, 7) = 71.73$, $p < .001$, and also that there were significant differences among the word-picture combinations, $F\ (3, 21) = 10.39$, $p < .001$. The Orientation × Displays interaction was also significant, $F\ (3, 21) = 20.50$, $p < .001$.

The interaction between face orientation and differences among word–picture displays appears contrary to the predictions of the linguistic model of Chase and Clark (1971). If inversion of the face added a constant to the duration of the picture-encoding stage, and picture encoding, word encoding, and comparison were truly independent operations, the effects for displays and face orientation would be independent. As it is, the results for the normal face appear similar to those obtained in the previous studies, in that they show quite large *above–below* and *true–false* differences, whereas these differences are absent, or in the opposite direction, when the face is inverted.

As a check on these conclusions, an analysis was carried out on the data for the normal face condition in which response (*yes* vs. *no*) and word (*above* vs. *below*) were factors. The differences of 163 msec. between the means for *yes* and *no*, and of 91 msec. between *above* and *below* displays, were significant, with $Fs\ (1, 7) = 35.74$, $p < .001$, in both cases, although the Response × Word interaction was not significant, $F\ (1, 7) < 1$. A similar analysis of RTs for the inverted face displays failed to show a significant response effect, $F\ (1, 7) < 1$. Displays with the word *below* were classified 60 msec. faster than displays with the word *above*, and this reversal of the usual effect was significant, $F\ (1, 7) = 6.93$, $p < .05$. The Response × Word interaction was not significant.

In the case of *yes* responses, inversion of the face resulted in a delay of response of about 200 msec. for *above*–above combinations, and about 90 msec. for *below*–below combinations. The difference between the 2 *same* displays was significant when the face was normally oriented, $F\ (1, 7) = 13.41$, $p < .025$, but not when the face was inverted, $F\ (1, 7) = 1.46$. Thus, for congruent combinations, inversion of the face resulted in an overall delay in RT, and also in elimination of the *above–below* difference. This outcome appears consistent with the predictions of the response availability theory but not those of the linguistic or perceptual theories. The linguistic theory predicts maintenance of an *above–below* difference when the face is inverted, whereas the perceptual theory predicts a reversal of this difference. The count of positive features shown in Table 3 indicates that the *above–below* effect should be attenuated when the face is inverted, and it was this result that was obtained.

Further analyses of *no* response times indicated that inverted face displays were classified 33 msec. faster than normal face displays, but that this difference was not significant, $F\ (1, 7) = 3.13$. The effect for words (*above* vs. *below*) also was not significant, but there was a significant Orientation × Words interaction, $F\ (1, 7) = 24.67$, $p < .01$. The absence of an effect for face orientation and the occurrence of the interaction appear counter to the predictions derived from the linguistic theory. The result is also in the wrong direction for the perceptual theory, since displays having the dot in the lower region of the figure were classified significantly faster than displays having the dot in the upper region, $F\ (1, 7) = 24.67$, $p < .01$. However, the interaction is consistent with the predictions of the response availability model. Summed negativity has a value of 4 for *above*–below combinations for both normal and inverted orientations of the face. When the face is normal, the representation of the *below*–above display contains only 3

negative components. Thus, under this condition, *no* responses should be faster to *above*–below displays than to *below*–above displays, and this is indeed the case, $F\ (1, 7) = 20.93$, $p < .01$. When the face is inverted, the sum of negative components in the *below*–above representation increases to 5. This increment in negativity is expected to lower the threshold for selection of a *no* response. The data indicate that, under this condition, *below*–above displays were classified significantly faster than *above*–below displays, $F\ (1, 7) = 12.63$, $p < .025$.

A further point of difficulty for both the perceptual and the linguistic theory is raised by the failure to obtain a *true–false* effect under the inverted face condition. If, as both theories imply, inversion of the face has its main effect on the duration of a picture-encoding stage, and this stage is independent of a comparison stage, there should be no effect on the magnitude of the *true–false* difference. In the model of Chase and Clark (1972), and Clark and Chase (1972), the *true–false* effect is defined as the time required to alter a "truth index" from "true" to "false." If this is correct, a *true–false* effect would be obtained in both the normal and the inverted face conditions. Thus, the absence of an effect in the present situation counts as evidence against Chase and Clark's (1971, 1972) assumptions about the independence and additivity of the encoding and comparison stages. The response availability model is sufficiently flexible to accommodate variations in the size of the *true–false* effect. For example, presence of a negative feature for face orientation might raise thresholds for *yes* responses and lower thresholds for *no* responses, thus overriding a general bias in favor of *yes* responses.

General Discussion

The analyses of these 2 experiments appear sufficient to eliminate the perceptual theory of Seymour (1969) as an explanation of the *above–below* asymmetry observed in word–picture comparison tasks. The theory must have difficulty in accommodating (*a*) the difference between *above*–above and *below*–below observed for horizontal face displays, (*b*) the absence of a difference between these 2 displays observed when the [illegible] inverted, (*c*) disappearance of the [illegible] effect when the face was inverted, and [illegible] tendency is Experiment II for *no* responses to occur faster to displays having [illegible] lower region than to displays having the dot in the upper region.

The experimental results also appear sufficient to disconfirm Chase and Clark's (1971 assumption that the RT in this type of comparison task is the sum of the [illegible] independent and successive w[illegible] picture encoding, comparison and res[illegible] stages, which differ in duration for *above* v[illegible] *below* concepts and for *true* vs. *false* displays If word encoding and picture encoding were truly independent operations, an alteration in the orientation of the reference face would affect only the duration of the picture-encoding stage. Further, if the comparison depended on establishment of abstract assertions of relative location, and above assertions were established quicker than below assertions, picture encoding should be faster for above displays than for below displays even when the face is inverted or rotated to the horizontal. The difference in time to encode the words *above* and *below* should also remain unaffected, as should the comparison time difference between *true* and *false* displays. The data are inconsistent with this position in the foll[illegible] respects: (*a*) The overall effects [illegible] (*above* vs. *below*) and picture locati[illegible] vs. below) were not significant when the [illegible] was horizontally oriented or when [illegible] inverted. (*b*) The *true–false* effe[illegible] nated when the face was invert[illegible] version of the face did not result in [illegible]verall delay of reaction for *no* responses in Experiment II, but did result in a reversal in the direction of the difference between *above*–below and *below*–above combinations.

It has been argued that the obtained results are consistent with a response availability model of the type which has been described. The main assumption of the model is that variations in RT in the sentence–picture comparison task reflect momentary adjustments in thresholds for selection of *yes* or *no* responses which occur as the encoding the display proceeds. Positive features in the representations of word and picture meaning, or positive outcomes of matching operations, tend to lower the threshold for a *yes* response. It is a facilitation effect of this sort that underlies

the *above–below* difference, rather than an absolute inequality in encoding time. This explanation may be extended to other pairs which involve an unmarked/marked, or +polar vs. −polar distinction, such as *large* vs. *small* (Seymour, 1971). A formulation of this type is sufficiently flexible to accommodate cases where the effect does not occur. For example, both the *above–below* and the *large–small* effects may be eliminated by instructing *S*s to report *no* to *same* displays, and *yes* to *different* displays. In the present experiments, an alteration in the orientation of the reference shape was sufficient to attenuate or eliminate the *above–below* effect. It was argued that this might occur if a shift from normal vertical orientation of the reference face had the effect of introducing negative components into the semantic representation of the picture. As shown in Tables 2 and 3, this tends to decrease the difference in summed positivity scores as between the *above*–above and *below*–below combinations, and this in turn will reduce the difference in *yes* response times.

An important feature of the response availability model is that it predicts that facilitation of *no* responses may occur when the number of negative components in the representation of the display is increased. The summed negativity scores in Tables 2 and 3 correctly predict (*a*) the absence of a difference between *above*–below and *below*–above displays when the face is horizontal, (*b*) similar RTs for *above*–below for normal and inverted face displays, (*c*) faster times for *below*–above when the face is inverted than when it is normally oriented, and (*d*) a reversal in the direction of the difference between *above*–below and *below*–above for normal vs. inverted face displays. The model can also accommodate interactions of display characteristics with the size of the *yes–no* difference if negativity in the semantic representation tends to equalize the thresholds for the *yes* and *no* responses.

The main contention of this article, therefore, is that RTs obtained in a simple word–picture comparison task may be quite well represented in terms of effects of semantic positivity/negativity on thresholds for *yes* and *no* responses. Some question arises about the generality of these conclusions, and in particular whether the difficulty encountered by the additive stages model in relation to the present, relatively simple, experimental situation necessarily implies a weakness in the more complex situations examined by Chase and Clark (1972) and Clark and Chase (1972). Krueger (1972) has presented evidence that the encoding and comparison stages are not additive in their effects on RT in the sentence–picture comparison task, in that effects attributable to sentence form (affirmative vs. negative) interact with effects supposedly arising during encoding (axis match effect). In general, the consequences of a negative in the sentence should be quite similar to those of inverting the face (negating the picture) in Experiment II of the present study, and should tend to eliminate or reverse the *true–false* effect. Thus, the Affirmative–Negative × True–False interaction obtained by Clark and Chase (1972) and other investigators may be handled by the response availability model, if it is assumed that a negative in the sentence causes a general rise in both response thresholds, and also tends to lower the *no* threshold relative to the *yes* threshold.

REFERENCES

Chase, W. G., & Clark, H. H. Semantics in the perception of verticality. *British Journal of Psychology*, 1971, **62**, 311–326.

Chase, W. G., & Clark, H. H. Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1972.

Clark, H. H., & Chase, W. G. On the process of comparing sentences against pictures. *Cognitive Psychology*, 1972, **3**, 472–517.

Krueger, L. E. Sentence–picture comparison: A test of additivity of processing time for feature matching and negation coding. *Journal of Experimental Psychology*, 1972, **95**, 275–284.

Leech, G. *Towards a semantic description of English*. London: Longmans, Green, 1969.

Schaeffer, B., & Wallace, R. The comparison of word meanings. *Journal of Experimental Psychology*, 1970, **86**, 144–152.

Seymour, P. H. K. Response latencies in judgements of spatial location. *British Journal of Psychology*, 1969, **60**, 31–39.

Seymour, P. H. K. Perceptual and judgemental bias in classification of word-shape displays. *Acta Psychologica*, 1971, **35**, 461–477.

Seymour, P. H. K. Judgements of verticality and response availability. *Bulletin of the Psychonomic Society*, 1973, **1**, 196–198.

(Received May 1, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 456–462

CONTEXT AND FREQUENCY EFFECTS IN THE GE[illegible]TION OF A HUMAN VOLUNTARY RESPONS[illegible]

JOHN A. HÉBERT,[1] MARSHA BULLOCK, LYNN LEVITT, KIM GR[illegible]
AND FRANK D. McGUIRK

Colorado State University

Three studies investigated the effect of presenting a given tes[illegible] more frequently than the other test stimuli on generalization. [illegible] training on a single stimulus, either 100, 150, or 200 gm., *Ss* wer[illegible] identify the stimulus in a test series of 5 weights, ranging in value 10[illegible] in 25-gm. steps. Systematically overrepresenting each of the 5 t[illegible] produced an ordered set of generalization gradients. A tenden[illegible] generalization responses to shift toward the overrepresented test stim[illegible] observed. The data were interpreted in terms of adaptation level t[illegible] Direct evidence was found for the link between generalization resp[illegible] adaptation level operationally defined by the category-rating method.

In human voluntary stimulus generalization, the context in which the original training stimulus (S+) is placed in testing has been shown to greatly alter the shape of the gradient of stimulus generalization. In an asymmetrical context (e.g., S+ is either the largest or smallest of the test values), *Ss* tend to choose central values of the test range as the original training stimulus more often than the S+ value itself (Thomas & Jones, 1962). This central-tendency effect has since been found with the following stimulus dimensions: size–area (Helson & Avant, 1967), weight (Hébert, Origlio, & McGuirk, 1972; Hébert & Capehart, 1969), brightness (Thomas, Svinicki, & Vogt, 1973), and line angle (Giurintano, 1972). However, Thomas and Jones, using hue as a dimension, also found that *Ss* respond accurately to S+ as the original training stimulus if S+ is placed symmetrically in the middle of the test values. That *Ss* tend to respond to central values of the test range *regardless* of the position of S+ in that test range was confirmed by Helson and Avant.

Related to this context effect is the finding that presenting S+ more frequently than the other test stimuli negates the central-tendency effect in an asymmetrical test context (Hébert & Capehart, 1969; Hébert et al., 1972). In both studi[illegible] following training on single weight [illegible] group of *Ss* was tested with an "overrepresentation" proced[illegible] [illegible]ich S+ was presented 5 times m[illegible]quently than other test stimuli. Another group was given an "equal-presentation" procedure in which S+ was presented equally as often as the other test stimuli. Since S+ was the smallest value in the test series, creating an asymmetrical context, a central-tendency effect would be expected. However, the overrepresentation-procedure group did not tend to choose central values as S+, as did the equal-presentation group, but rather, correctly chose the S+ value more often than any other test stimulus, yielding a downward, linear gradi[illegible] Thus, a strong frequenc[illegible] with a context effect. Giurintano (1972) has also demonstrated [illegible]nomenon with a line angle dimensi[illegible]

Capehart, Tempone, and Hébert (1969) interpreted these data in terms of an adaptation level (AL) approach to a stimulus generalization. According to AL theory, when *S* is presented with a single stimulus value in training, a training AL is established at that value on a subjective dimension. What *S* learns is not "stimulus of value *x* is correct" but "the stimulus whose value corresponds to AL is correct." Thus, if in testing AL shifts away from S+ for some reason such as an asymmetrical context, then a concomitant shift in *S*'s

[1] Requests for reprints should be sent to John A. Hébert, Department of Psychology, Colorado State University, Fort Collins, Colorado 80521.

judgment of which stimulus is S+ will be observed. The AL by which *S* makes a judgment can be represented as a weighted average of the training and test ALs:

$$\text{AL} = w_1\text{AL}_{\text{train}} + w_2\text{AL}_{\text{test}}, \qquad [1]$$

where w_1 and w_2 are subjective weights which sum to 1 and reflect the relative contributions of training and testing to the prevailing AL. With this model, varying the *amount* of training or of testing would alter the values of the subjective weights and thus result in different prevailing ALs. Similarly, the value of the training AL can be manipulated by choosing different values for S+, while the value of the test AL can be manipulated by choosing different test stimuli. In addition, the value of the test AL can be altered by manipulating the frequency of presentation of test stimuli. For example, when S+ is overrepresented in testing, greater weight is given to it in averaging the stimuli to obtain AL (Capehart et al., 1969; Helson, 1964). The greater weight given the overrepresented stimulus biases or "pulls" AL toward that value.

However, the frequencies of stimuli other than S+ can also be manipulated. Suppose, for example, that S+ is the smallest of 5 test stimuli, and that the stimulus value *most distant* from S+, say S_5, is presented with much greater frequency than the other 4 test stimuli, including S+. Under these circumstances, S_5 would have greater weight, pulling AL towards it, thus producing a larger discrepancy between training and test ALs. This would produce the greatest amount of shift in the gradient, possibly *beyond* the center of the test values. A design that systematically manipulates the frequencies of presentation of each test stimulus independently could provide a more exacting test of the AL approach. An ordered set of gradients should be obtained if the model is at least correct in an ordinal sense.

Anderson's (1971) work in integration theory and its relation to AL (Anderson, 1973) also suggests a rigorous test of the averaging hypothesis of Equation 1. A factorial design in which both the training and test ALs are manipulated should yield test data which must meet the key prediction of a simple averaging model: A graphical representation of the interaction between training and testing manipulation should yield parallel curves. That is, the Training × Testing interaction should be "zero in principle and nonsignificant in practice [Anderson, 1970, p. 155]." What is needed to make this test of the averaging model is a response measure directly related to the prevailing AL.

The present experiments were designed to test the effects of systematically manipulating the frequencies of presentation of the test stimuli in both symmetrical and asymmetrical test contexts. In addition, an experiment to tie *S*'s responses directly to Helson's (1947, 1964) operationally defined AL concept was conducted.

Method

Experiments I and II were designed to test the effect of manipulating test stimulus frequencies following training on either the lightest or heaviest of the test stimuli (Experiment I) or following training on the intermediate value of the test stimuli (Experiment II).

Subjects and apparatus. A total of 120 students served as *S*s in Experiment I; 60 students served as *S*s in Experiment II. The *S*s in both experiments were participating as a course requirement. The stimuli consisted of a set of 5 weights ranging in value 100–200 mg. in 25-gm. steps. The weights were made from aluminum cans 10.8 cm. tall and 5.5 cm. diameter, and were filled with wax and shot and painted to appear identical. When the blindfolded *S* lifted a given weight with the preferred hand, a microswitch was released which started a timer. When *S* reached a decision (*same* or *different*), he pressed 1 of 2 response keys with the index finger of his free hand, which rested above and between 2 closely spaced response keys, labeled same and different. The pressing of 1 of the 2 keys stopped the timer and activated a response light, and *E* then recorded choice and latency of response.

Procedure. The *S*s were told that they were participating in a weight perception experiment. In Experiment I, they were instructed to lift the training weight or S+ (either 100 or 200 gm.) 5 times and to remember it because they were to identify it in a test series of weights immediately following training. They were told to press a *same* key if a test stimulus was the same as S+, or to press a *different* key if they felt it was different (either heavier or lighter). In addition, *S*s were asked to respond "as quickly and accurately as possible." Following instructions, all *S*s were

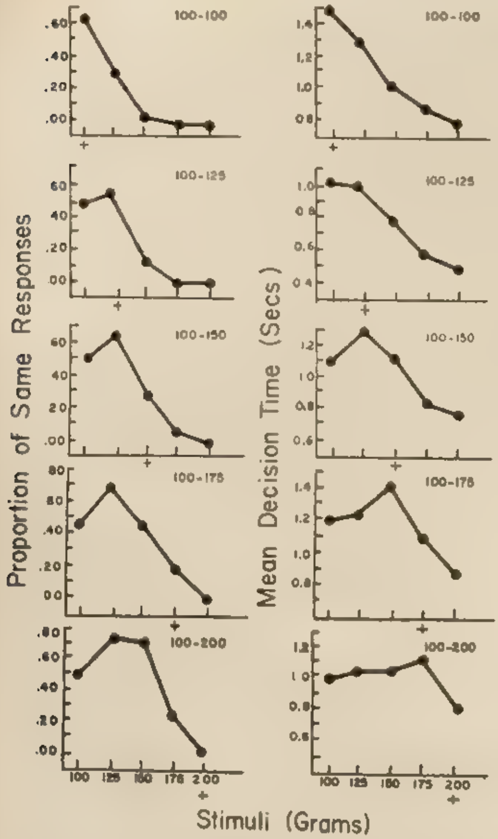


FIGURE 1. Generalization gradients of *same* responses and decision time for the 100-gm. training group.

blindfolded for the rest of the session. The training procedure for Experiment II, conducted after the completion of Experiment I, differed in only one respect, namely, *S*s were trained on 150 gm., the intermediate value of the test series. Testing procedures were identical in both experiments.

Following training, which consisted of lifting S+ 5 times in a slow, paced fashion, test trials were immediately begun. A test trial consisted of *S*'s grasping the weight from the top with 4 fingers and a thumb, lifting the weight up with elbow resting on table, making the response, bringing the weight down, and releasing it. Shaking or swinging the weight was not permitted.

In Experiment I, half of the *S*s were trained on 100 gm. and half on 200 gm. Following this, *S*s in both groups were assigned randomly to 1 of 6 test groups, 10 *S*s per group, yielding a 2 × 6 factorial design. The 6 test groups for both the 100- and 200-gm. training groups consisted of a control group in which the 5 test stimuli were presented equally often (5 times each) and 5 other groups which differed in terms of which test stimulus was overrepresented. There were 25 presentations of the overrepresented stim[illegible] [illegible]presentations of each of the other stimuli. [illegible] procedure was identical in Experiment I[illegible] [illegible] groups of 10 *S*s each with 1 group an [illegible]presentation control group and the remaining [illegible] differing [illegible] of which stimulus was [illegible]represented. For discussion purposes, groups [illegible] identified by designating first the traini[illegible] (100, 150, and 200 gm.) and then the st[illegible] represented for that group. For exam[illegible] 100-175 was trained on 100 gm. and in [illegible] presented 5 times more frequently tha[illegible] stimuli. The equal-presentation control [illegible] labeled 100-E and 200-E for Experiment I, [illegible] for Experiment II.

Stimuli were presented in a [illegible]mized order within a block of 9 trials, wit[illegible] the overrepresented sti[illegible] 3 times in succession. A block of 9 trials consisted of 5 presentations of the overrepresented stimul[illegible] intermixed with 1 presen[illegible] of the other remaining stimuli. In [illegible] of trials, a total of 45 trial[illegible] the 100-E and 200-E co[illegible] I and the 150-E group f[illegible] were given 5 blocks of 5 [illegible]

Design and analysis. Th[illegible] Experiment I were analyzed separately fr[illegible] those of Experiment II. Analysis for Experiment I involved a 2 × 5 × 5 factorial design with repeated measures, with 2 training conditions and the 5 overrepresentation groups as between-subjects variables, and the 5 test stimuli as the repeated variable. Two dependent measures were analyzed by this design, proportion of *same* responses and response latency. The proportion of *same* responses to all response[illegible] by a given *S* across blocks of trials served as [illegible] main dependent measure. Latency, defined [illegible] time from lift to key press, was averaged [illegible] trials for each stimulus and *S*. These late[illegible] were calculated on the basis of *both same* and *different* responses to a given stimul[illegible] Data from the 100-E and 200-E control gr[illegible] separately and compared [illegible] group of Experiment II. [illegible] ment II involved a 5 × 5 [illegible] with the 5 overrepresentation grou[illegible] test stimuli as a repeated variable.

RESULTS AND DISCUSSION

The generalization gradients for each of the 10 experimental groups are depicted in Figures 1 and 2 for the 100- and 200-gm. training groups, respectively. The left-hand columns of both figures depict gradients of *same* responses for each group, while the right-hand columns represent the latency data for the same groups. Inspection of the choice gradients in the left-hand columns of Figures 1 and 2 indicates

a systematically increasing tendency to choose stimuli farther from S+ as the overrepresented stimulus becomes farther from S+, as predicted from the AL model. Statistically, this is represented by a significant Test Procedure × Stimuli interaction, F (16, 360) = 5.13, $p < .001$. However, the amount of shift appears greater for the 200-gm. training groups, resulting in a significant Training × Test Procedure × Stimuli interaction, F (16, 360) = 4.31, $p < .001$. The most dramatic shift occurred with Group 200–100, whose data are presented in the bottom left-hand panel of Figure 2. This group shifted *past* the center (150 gm.) to some point between 100 and 125 gm. In contrast, Group 100–200, the counterpart of Group 200–100, failed to shift even quite to the center (see lower left-hand panel of Figure 1), if shift is defined by the modal response value or peak. This asymmetry may reflect the fact that the *perceived* midpoint of these stimulus values is closer to 125 gm. than to 150 gm., the arithmetic midpoint, when perceived midpoint is defined by AL. The AL of this test series is calculated to be approximately 128 gm. according to Helson's (1947, 1964) formula for lifted weights, and this value is confirmed by category-rating data from our laboratory. The notion of "central-tendency effect" (Thomas & Jones, 1962) might best be conceptualized as a tendency to respond to the psychological center, not the physical center.

Inspection of Figures 1 and 2 indicates a correspondence between choice and latency data. The stimuli most often associated with *same* responses are also associated with the longer latencies than those of stimuli rarely chosen *same*, though this correspondence is far from exact. Capehart et al. (1969) offered the explanation that stimuli close to AL are less "distinctive" than those far from AL, and therefore, take longer to respond to than the more "distinctive" stimuli far from AL. Since *same* responses are associated with AL, then *same* responses take longer.

The gradients of *same* responses of the 5 experimental groups from Experiment II

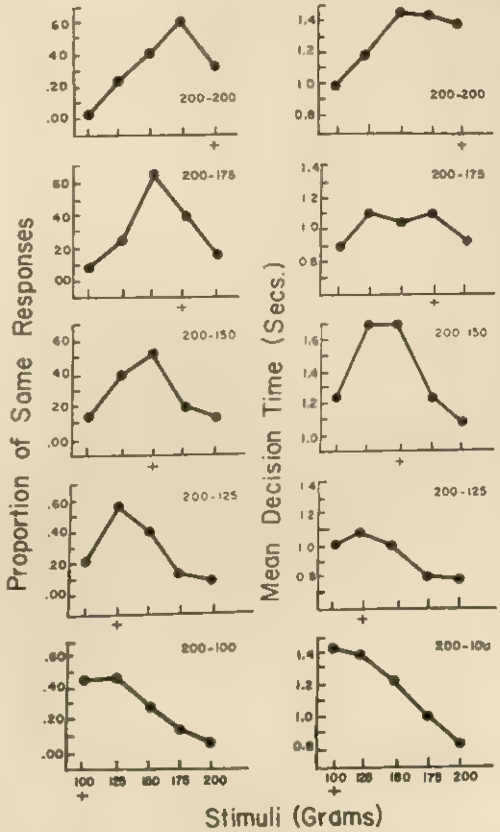


FIGURE 2. Generalization gradients of *same* responses and decision time for the 200-gm. training group.

are depicted in Figure 3A. It is clear from the inspection of these gradients that the same systematic effects of the frequency manipulation occurred here as in Experiment I, resulting in a significant Test Procedure × Stimuli interaction, F (16, 180) = 3.56, $p < .001$. The most notable aspect of Figure 3A is the asymmetry: The modal response value is 125 gm., not 150 gm., and the 100-gm. stimulus receives many more responses than its counterpart 200 gm. This again may be due to the fact that perceived midpoint of this series (defined by AL) is close to the 125-gm. stimulus, not 150 gm. Figure 3B depicts the gradients of the 100-E, 150-E, and 200-E control groups from both experiments, and again, the tendency to a shift in response toward the 125-gm. stimuli is apparent. The term *central tendency* should

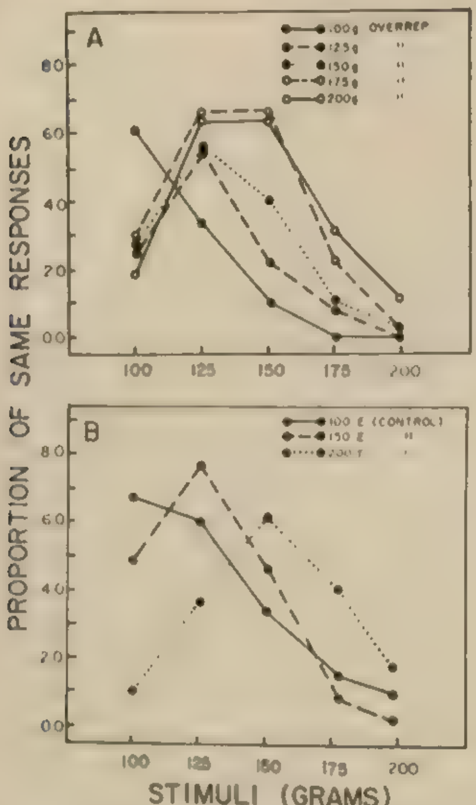


FIGURE 3. Generalization gradients for the experimental groups trained on the 150-gm. weight are plotted in the upper panel; generalization gradients for the equal-presentation control groups for Experiments I and II are plotted in the lower panel.

be used with caution since the 150-E group shifted *away* from the center of the series, although one could say *toward* AL, the psychological center. Note that the 100-E group did not actually peak at 125 gm., as in the Hébert and Capehart (1969) study, but at 100 gm. However, compared to the 100-100 group, there was a much greater tendency to respond to central values. An analysis comparing just these 2 groups still yields a significant Test Procedure × Stimuli interaction, F (4, 72) = 3.22, $p < .05$.

While the above data are consistent with the AL approach, a rigorous test of the model as expressed in Equation 1 is also possible. Normally, AL is measured by means of a category-rating method (Helson, 1947), but in this situation no ratings were taken. Capehart et al. (1969) proposed that the modal or peak response value is closest to AL in generalization experiments, and as such, is a crude indicator of AL. Unfortunately, this measure ignores responses to all other stimuli, and is, therefore, biased. A better index is the *mean response value* (Thomas, 1962) which takes into account responses to all stimuli by weighting each stimulus value by the frequency with which it was called *same*. The mean response value for each S in the 15 experimental groups in Experiments I and II was calculated and analyzed by means of a 3 × 5 completely randomized factorial design, with 3 training conditions (100, 150, and 200 gm.) and 5 test procedures (overrepresentation of 100 through 200 gm.). In terms of the additive model of Equation 1, 3 training ALs are factorially combined with 5 test ALs, and, further, a plot of the curves of these groups should be parallel, or noninteractive. The Training × Test Procedure interaction depicted in Figure 4 was not statistically significant, F (8, 135) = .65, $p > .05$, indicating that no meaningful interaction is present. However, Anderson (1973) has indicated that nonparallelism should be present in the form of a bilinear interaction. Since

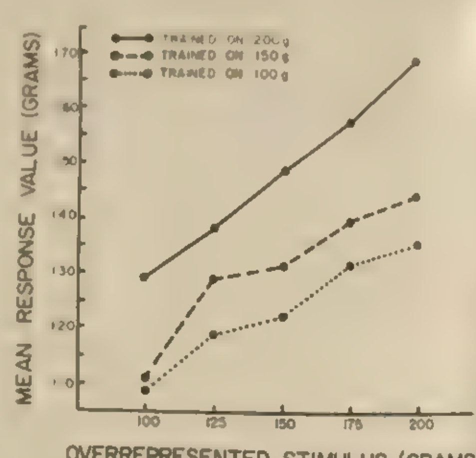


FIGURE 4. The nonsignificant Training × Test Procedure interaction for the 15 independent groups from Experiments I and II.

this interaction is concentrated in only 1 degree of freedom, a specific bilinear F test was made, but no significant interaction was found, $F\ (1,\ 135) = 2.27$, $p > .05$. Thus, the additive model of Equation 1 is not disconfirmed and remains a tenable hypothesis.

The question still remains whether the concept of AL as used in this study is substantially the same as the concept as operationally defined by Helson (1947, 1964); neither is there any direct evidence that *S*s were actually linking *same* responses to AL. A third experiment was conducted to answer these questions. Nine introductory psychology students served as *S*s. Training and testing of these students was identical to the 150-E control group. That is, *S*s were asked to heft the 150-gm. training weight 5 times, and then were required to identify S+ in a subsequent test series of 5 presentations each of the 5 test stimuli. However, the testing procedure differed in one important respect: Following each *same* or *different* response each *S* was asked to rate the weight just lifted on a 7-point category-rating scale (1 = very light, 2 = light, 3 = fairly light, 4 = neither light nor heavy, 5 = fairly heavy, 6 = heavy, 7 = very heavy). The 150-gm. training stimulus was not identified at any point on the rating scale by *E*, and no instructions to do so were given to *S*.

With this rating procedure, the stimulus which is closest to an average 4 rating can be assumed to be closest to AL, which would be operationally equivalent to a stimulus rated exactly 4. Figure 5 depicts the patterns of choice and latency of response for this third experiment. Inspection of the gradient of *same* responses (solid line shows that 125 gm. elicited the greatest proportion of *same* responses, not the 150-gm. S+). This is in close agreement with the gradient from Group 150-E shown in Figure 3B. As predicted, the 125-gm. stimulus also had the mean rating closest to 4.00 (100 gm. = 2.62, 125 gm. = 3.87, 150 gm. = 4.51, 175 gm. = 5.40, and 200 gm. = 6.20). The most compelling evidence that AL and *same* responses

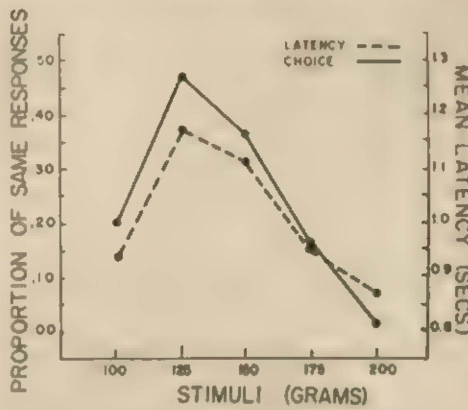


FIGURE 5. The latency and choice gradients of generalization for Experiment III.

are linked is the finding that 50 of the 54 *same* responses in this study were to stimuli which were subsequently rated 4. Additionally, none of the *different* responses were given a 4 rating.

These 3 studies taken together cast doubt on the assumption that physical similarity is the primary basis of stimulus generalization of voluntary responses in human *S*s. The powerful effects of frequency and context must be accounted for in any theory of voluntary stimulus generalization. The Capehart et al. (1969) theory of stimulus equivalence provides such an account, and the data in these studies offer strong support for the theory. As a final note, the reader should be reminded of the possibility of interpreting these data with Parducci's (1965) range-frequency model. Since both the AL and range-frequency approaches have generally the same qualitative predictions, range-frequency theory, with additional assumptions relating to a learning context (e.g., What is learned?), may be useful for the interpretation of these data.

REFERENCES

ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, 77, 153-170.

ANDERSON, N. H. Integration theory and attitude change. *Psychological Review*, 1971, 78, 171-206.

ANDERSON, N. H. Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.),

Handbook of perception. Vol. 2. New York: Academic Press, 1973.

Capehart, J., Tempone, V. J., & Hébert, J. A theory of stimulus equivalence. *Psychological Review*, 1969, **76**, 405–418.

Giurintano, L. P. Stimulus generalization as a function of training and test stimulus parameters. Unpublished doctoral dissertation, University of Colorado, 1972.

Hébert, J. A., & Capehart, J. E. Generalization of a voluntary response as a function of presentation frequency of the training stimulus in testing. *Psychonomic Science*, 1969, **16**, 315–316.

Hébert, J. A., Origlio, D. P., & McGuirk, F. D. Training and testing effects in the generalization of a voluntary response. *Psychonomic Science*, 1972, **26**, 209–210.

Helson, H. Adaptation-level as a frame of reference for prediction of psychophysical data. *American Journal of Psychology*, 1947, **60**, 1–29.

Helson, H. *Adaptation level theory.* New York: Harper & Row, 1964.

Helson, H., & Avant, L. L. Stimulus generalization as a function of contextual stimuli. *Journal of Experimental Psychology*, 1967, **73**, 565–567.

Parducci, A. Category judgment: A range-frequency model. *Psychological Review*, 1965, **72**, 407–418.

Thomas, D. R. The effects of drive and discrimination training on stimulus generalization. *Journal of Experimental Psychology*, 1962, **64**, 24

Thomas, D. R., & Jones, C. G. Stimulus generalization as a function of the frame of reference. *Journal of Experimental Psychology*, 1962, **64**, 77–80.

Thomas, D. R., Svinicki, M. D., & Vogt, J. Adaptation level as a factor in human discrimination learning and stimulus generalization. *Journal of Experimental Psychology*, 1973, **97**, 210–219.

(Received June 2, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 463–472

RELATIONSHIPS AMONG HIGHER ORDER ORGANIZATIONAL MEASURES AND FREE RECALL[1]

JAMES W. PELLEGRINO[2] AND WILLIAM F. BATTIG

University of Colorado

Higher order subjective organization units were measured in categorical and unrelated lists under fixed and random presentation schedules. Such units occurred frequently and also varied in sequential consistency across recalls thereby demonstrating the inadequacy of pair-wise sequential indices for measuring total organization. Further data confirmed the necessity for examining such units since, when included as additional predictors, they accounted for 96% of the between-groups recall variance. Several other organization measures indicated interactive relationships between organizational processes and changes in such relationships with practice which may partially explain inconsistencies in the organization-recall literature. Additionally, the present organizational data raise questions about the adequacy of current theories of storage and retrieval processes in free recall.

Recent research in free recall has focused on the role of organizational processes in memory and questions concerning the storage and retrieval of information as related to such processes (e.g., Postman, 1972; Slamecka, 1968). Postman has noted 2 major current problems related to the study of organizational processes in free recall.

The first problem concerns the adequacy of certain organizational measures as valid indices of the total organization that has occurred at any point in learning, particularly with respect to measures of subjective organization (SO). The 2 major indices of SO—the intertrial repetition measure (Bousfield & Bousfield, 1966) and the SO measure (Tulving, 1962)—have been criticized as having very serious shortcomings both in the size of the sequential units that are examined and the strict sequential criteria that are employed. Shuell (1969) and others (e.g., Postman, 1972) maintain that such simple pair-wise sequential indices would seriously underestimate the actual organization that occurred if *S*s formed units containing more than 2 items with free-varying internal sequential constraints, i.e., the items within a unit always appear together but not necessarily in the exact same sequential order each time. Specifically, a unit of Size 5 could be randomly reproduced 5! possible ways which yield an average of only 1.6 pair-wise repetition as compared to a maximum of 4. Empirical justification for such criticism of the current measures therefore requires demonstrating (*a*) the existence of such higher order units and (*b*) that such units have internal sequential variability across successive recalls.

Pellegrino (1971) has presented a technique for measuring organization which is capable of dealing with both issues. Under this procedure, the degree of sequential consistency can be examined for units of any size rather than being limited only to units of Size 2. This technique also permits measurement of organization based on any of the following 3 different sequential criteria: (*a*) *unidirectional*, requiring exact reproduction of the sequen-

[1] This article is Publication No. 37 of the Institute for the Study of Intellectual Behavior, University of Colorado. The research was supported by United States Public Health Service Fellowship MS-8030-0151 to the first author from the National Institute of Mental Health, and National Science Foundation Research Grant GB-25433 to the second author. The authors gratefully acknowledge the assistance of Shu-in Huang, Jerry Peebles, and Clara Chapman in collecting and tabulating the data, and of the Computer Laboratory for Instruction in Psychological Research for making possible the organizational analyses.

[2] Requests for reprints should be sent to James W. Pellegrino, Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213.

tial order, as characterizes most previous measures; (*b*) *bidirectional*, counting also a mirror-image reversal as an organizational unit; and (*c*) *unordered*, requiring only that the items be recalled together in any sequential order. For example, if Items A, B, and C were recalled in that order on Trial t, on Trial $t+1$ only A–B–C would represent a unidirectional unit, either A–B–C or C–B–A would be bidirectional units, whereas any of the 3! possible orderings of these 3 items would constitute unordered units. In the present measure the actual number of repetitions of each type is separately determined for Unit Sizes 2–5 and compared to an appropriate chance expected value. The actual chance deviation (observed–expected) is divided by the maximum possible deviation (maximum–expected) to provide a ratio organization score (adjusted ratio clustering, ARC′) whereby 0 represents chance and 1.0 perfect organization.

The initial purpose of this experiment was to employ this technique to examine how extensively higher order units occur, their sequential properties, and the effects of certain free-recall variables on the development of such units. To demonstrate the inadequacy of pair-wise measures requires not only the existence of higher order units, but more importantly they must be shown to vary in internal structure from trial to trial. If there are no differences between the 3 sequential ordering criteria, then pair-wise sequential measures would appear adequately sensitive.

A second major problem concerns the relationship of free-recall acquisition performance (mean recall) to the amount of organization. Several different types of organizational processes have been shown to function in free recall, e.g., primacy and recency effects (Tulving, 1968), priority of recall of newly learned items (Battig, Allen, & Jensen, 1965), taxonomic clustering (Bousfield, 1953), SO (Tulving, 1962), and correspondence of recall order with list presentation order (input–output organization or seriation, Mandler, 1969). Investigations of the nature of any relationships between organization and recall have been severely handicapped by their reliance upon only one type of organizational measure and the failure to consider other viable types of organization or changes in organization occurring with practice. Thus the inconsistencies in this literature may be a function of a complex interactive relationship between several different modes of organization occurring simultaneously, any one of which is capable of producing recall differences. It might also be the case that increases in one type of organization occur at the expense of other types of organization. Thus the present study also provides a detailed examination and comparison of all the aforementioned types of organization, in order to evaluate their interactive nature, their relative importance as a function of different experimental variables, and any changes over trials in the usage of various processes. Also examined is the relationship(s) of these various types of organization to recall performance.

Two of the 3 variables in this stu (categorical vs. unrelated word lists a fixed vs. random presentation orders) have been previously examined with respect to SO. Puff (1970) failed to find any differences between categorical and unrela word list when measuring pair-wise S The list-order variable has been examined in lists of related and unrelated materials (e.g., Winograd, Conn, & Rand, 1971). Both the acquisition and organizational data indicate a facilitative effect produced by a fixed serial order for unrelated words, but little or no effect for related materials. It was hypothesized that both the list structure and the list-order variable should produce large effects on SO and therefore provide the necessary test of the existence and sequential consistency of higher order units. The third variable, not previously examined in free-recall tasks, involved a categorization pretest procedure given to half of the *S*s. Developed by Loftus and Scheff (1971), this pretest provided *S*s with words later used in the actual free-recall list to which *S*s produced the labels of categories which each word represented.

Method

Design and materials. The design of this experiment was a 2 × 2 × 2 factorial manipulating list structure (taxonomic or unrelated), presentation order (fixed or random), and categorization pretest (pretest or no pretest). The 25-item taxonomic list consisted of 5 items from each of 5 taxonomic categories (animals, clergy, metals, instruments, parts of the body), with all items high in taxonomic frequency (Battig & Montague, 1969). The 25-item unrelated list consisted of high taxonomic frequency items chosen from 25 different categories. In both lists there were the same 16 initial letters, and the Kucera and Francis (1967) word frequency was also equal. There were 6 different orders of each list with 5 used for the learning task and the other used for the pretest and posttest. In the taxonomic list, no 2 items from the same taxonomic category ever appeared in adjacent sequential positions, and in both lists each word appeared in each fifth of the list across the 5 random orders used for learning. For the random order conditions, each of the 5 different list orders was used equally often as the first-trial order, thus counterbalancing the random list order sequences in a 5 × 5 Latin square. Within the fixed-order conditions, each of the 5 different list orders was also used equally often for different *S*s, thus equating fixed and random conditions with respect to frequency of occurrence of each of the 5 list orders.

Apparatus and procedure. All *S*s were seated at a table and presented a test booklet containing 12 blank pages. They were instructed that they would see a list of words, one at a time at a 3-sec. rate, and that when they saw the words START RECALL they were to turn to the next blank page in their booklet and write as many words from the list as they could in any order they wished. They were further instructed that they had 1½ min. to do so and this procedure would continue for several trials. The list acquisition lasted for 12 trials. Under pretest conditions, *S*s were first taken to a different room and instructed that they would hear a list of words one at a time at a 10-sec. rate. They were asked to write on each new page of the small booklet provided as many category labels as they thought were appropriate to that word. Immediately following the pretest, these *S*s were taken into the experimental room and participated in the experimental task. Following the 12 free-recall trials, all *S*s were given a posttest identical to that given in the pretest conditions.

Subjects. There were 120 *S*s with 15 in each of the 8 conditions. All *S*s were University of Colorado students fulfilling a course requirement for experimental participation. Each *S* was tested for 50–60 min. in groups of 1–3. These *S*s were assigned to conditions using a block randomization procedure which assured that within each block of 8 test sessions each condition appeared once, and that within the first 40 test sessions each of the 5 counterbalancing orders for each condition appeared once. Following the first 40 test sessions, *S*s were assigned to conditions to complete the counterbalancing of list orders for those conditions where less than 3 *S*s had served in the initial session.

There were 4 *E*s, each testing approximately the same number of *S*s in each condition. There was a loss of 10 *S*s due to equipment problems or failure to follow instructions.

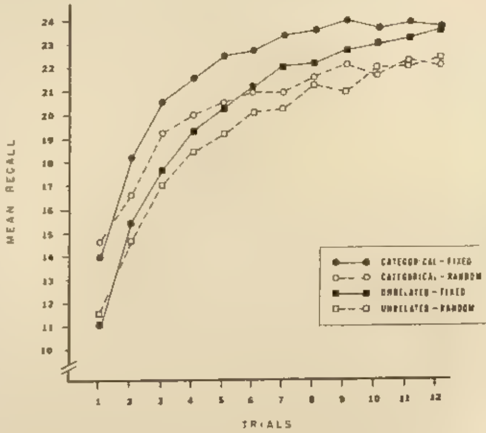


FIGURE 1. Mean recall as a function of list structure, list order, and trials.

Results

The minimum significance level for the mean recall and various organizational analyses that follow was .01. All main effects or interactions not specifically discussed in any of these sections did not reach this criterion. In each section that follows, the results are both described and related to previous empirical findings.

Acquisition. The basic recall results are presented in Figure 1. Mean recall was higher for categorical than unrelated list conditions, $F\ (1, 112) = 9.86$, and for fixed- than random-order conditions, $F\ (1, 112) = 8.62$. These differences decreased and increased, respectively, with trials, *F*s $(11, 1232) = 6.90$ and 3.31. No other interactions involving either variable approached significance. A large Pretest × Trials interaction, $F\ (11, 1232) = 10.00$, was primarily due to superiority of pretest over no-pretest conditions on the first 2 trials (15.76 vs. 13.25) and no effect thereafter (21.43 vs. 21.56), with this comparison accounting for 80% of the interaction variance.

TABLE 1

Mean Output-Output and Input-Output ARC′ Scores as a Function of Unidirectional and Unordered Criteria, Unit Size, and Point in Learning

Trials	Output-output			Input-output		
	Unidirectional	Unordered	Difference	Unidirectional	Unordered	Difference
Unit Size 2						
First half	.178	.265	.087	.069	.079	.010
Second half	.267	.372	.105	.105	.118	.013
Unit Size 3						
First half	.073	.159	.086	.038	.049	.011
Second half	.132	.258	.126	.085	.105	.020
Unit Size 4						
First half	.038	.105	.067	.026	.031	.005
Second half	.104	.208	.104	.073	.090	.017
Unit Size 5						
First half	.026	.077	.051	.020	.024	.004
Second half	.08[illegible]	.168	.084	.069	.081	.01[illegible]

These recall results are in general agreement with previous investigations (e.g., Mandler, 1969; Winograd et al., 1971). The one discrepant result was the facilitation by a fixed serial order within categorical lists. Mandler failed to find such facilitation, but obvious ceiling effects occurred in his 16-item lists. The failure of Winograd et al. to find such an effect may be due to the relatively short number of trials (4) since the present data indicate that this facilitation only occurs in later stages of learning. Thus the present results appear reconcilable with previous findings.

Ordering criterion effects. Initial analyses of both output-output and input-output organization ARC′ [illegible] scores [illegible] [illegible] 3 experimental variables, trials, and the 3 ordering criteria, and were performed separately for Unit Sizes 2–5.[3] This section is concerned with differences between the unidirectional and unordered criteria in sensitivity as organizational measures.[4]

Table 1 summarizes ordering criterion differences for both types of organization separately for each unit size. The values in Table 1 are pooled across all experimental conditions since all showed a similar pattern of results and any interactions

[3] Unit sizes were examined separately due to differences in the range of ARC′ scores that actually could be obtained. Although the theoretical range for ARC′ scores is identical for all unit sizes, any deviation from perfect sequential ordering (upon which the maximum observed values are based) produces larger reductions in the range of actual possible observable repetitions for the larger unit sizes, thereby reducing the functional range of ARC′ scores more severely for higher order unit sizes. For example, a single departure from perfect sequential consistency of 11 recalled items can produce ARC′ scores for units of Size 5 ranging .86–.29, while for units of Size 2 this would range only .90–.80 (depending on how close the discrepant item is to either end of the sequence). Any comparison across unit sizes involving the absolute numeric ARC′ score would therefore be misleading and to avoid such problems in interpretation separate analyses were performed on each unit size.

[4] Only unidirectional and unordered values are compared in this analysis due to the necessary identity of bidirectional and unordered scores for Unit Size 2, and data clearly showing that unidirectional and bidirectional values did not differ for Unit Sizes 3–5.

involving experimental conditions only reflected differences in the magnitude of the ordering criterion effect. The present discussion focuses only on the overall ordering criterion differences and a subsequent analysis will deal with the experimental conditions. The output–output analysis showed highly significant differences between criteria for all 4 unit sizes, $Fs\ (1, 112) > 206.65$. All 4 unit sizes also produced Ordering Criterion × Trials interactions, $Fs\ (10, 1120) > 3.56$. A subsequent breakdown of these interactions into first and second half of learning (as shown in Table 1) showed that the unidirectional–unordered difference was greatest in the second half of learning, particularly for the larger unit sizes, yielding $Fs\ (1, 1120)$ of 5.87, 8.98, 29.30, and 38.42 for Unit Sizes 2–5, respectively. In the input–output organization analyses, significant ordering criterion differences also occurred, $Fs\ (1, 112) > 23.42$, but these did not interact significantly with trials.

These ordering criterion effects indicate quite strongly that the unordered criterion is the most sensitive for determining the total amount of organization based upon either the previous recall order or the present list order. They further indicate that when *S* is using prior recall as a basis for determining present recall order, such organizational units vary considerably in sequential consistency, particularly for larger unit sizes late in learning. Thus simple pair-wise sequential measures necessarily underestimate the total amount of organization, so units larger than Size 2 must be examined to determine the effects of experimental variables upon SO. When *S* is basing recall order on the list presentation order, however, organizational units vary less in internal sequential consistency and this does not change over trials.

As noted previously, the ordering criterion difference also interacted with experimental conditions and these interactions generally showed that any differences between experimental conditions were primarily a function of the unordered values, particularly in the larger unit-size analyses. Since the unordered criterion more adequately represents and is more sensitive to the total amount of organization that has occurred, the analyses that follow are based upon only the unordered criterion scores.

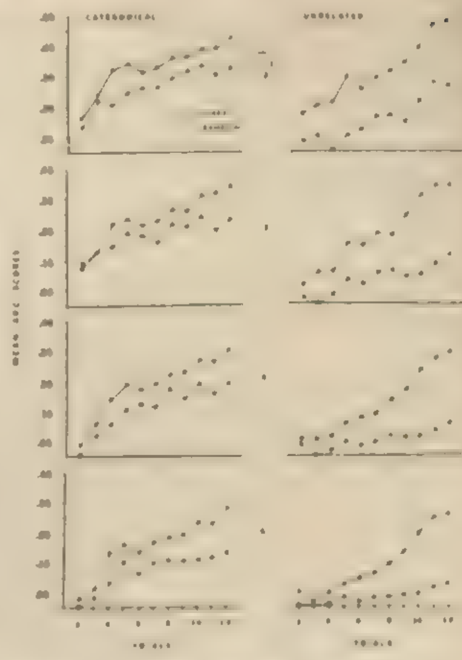


FIGURE 2. Mean unordered output–output organization for Unit Sizes 2–5 as a function of list structure, list order, and trials.

Output–output. The major results of these analyses are shown in Figure 2. The pretest variable had no effect in this or any other organizational analyses that follow, and will not be discussed further in those sections. Figure 2 shows overall significant increases in organization over trials for all 4 unit sizes, $Fs\ (10, 1120) > 24.25$. Differences between categorical and unrelated list conditions were found only in the analyses of Unit Sizes 3, 4, and 5, $Fs\ (1, 112) > 8.91$. Interactions between this list structure variable and trials also appeared in the analyses for Unit Sizes 2, 4, and 5, $Fs\ (10, 1120) > 2.60$, reflecting maximal differences between categorical and unrelated conditions on Trials 4–10 with little difference on the first and last 2 trials. Figure 2 also shows large differences between fixed- and random-list-

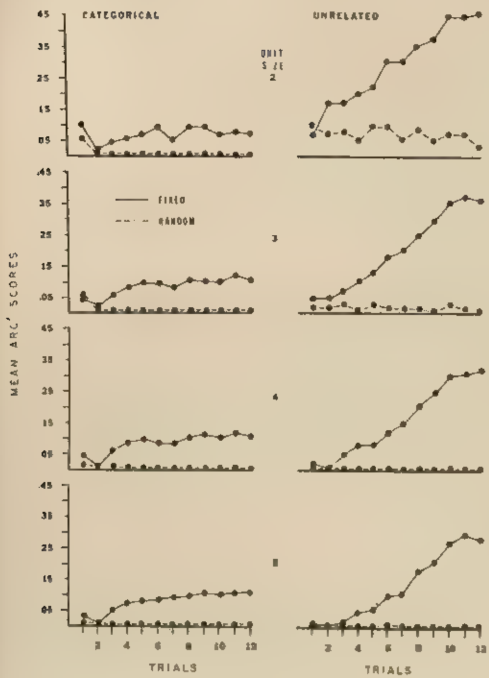


FIGURE 3. Mean unordered input-output organization for Unit Sizes 2–5 as a function of list structure, list order, and trials.

order conditions for all 4 unit sizes, *F*s (1, 112) > 13.45. However, List Order × Trials interactions only appeared in analyses for Unit Sizes 3, 4, and 5, *F*s (10, 1120) > 3.57, indicating increasing differences favoring fixed- over random-order conditions as learning progressed.

The present results for Unit Size 2 are consistent with previous data (e.g., Mandler, 1969; Puff, 1970). However, Puff's conclusion that the degree of SO is identical for randomized categorical and unrelated lists is contradicted by larger unit-size data. This result further indicates that pair-wise indices are relatively insensitive to the total amount of SO, particularly if that organization has a free-varying structure. Moreover, higher order unit-size measures produce differences more closely paralleling those for list acquisition, notably superiority of categoric over unrelated lists, and fixed over random orders especially on later trials. The latter results further indicate that higher order measures should be more closely related to between-groups recall differences, which would provide a strong argument for their usage. This was examined by computing separate correlations for recall with Unit Sizes 2 and 2 5. Pair-wise and multiple correlations based upon group means yielded respective values of .67 ($p > .05$) and .98 ($p < .001$). These correlations differed significantly, F (3, 3) = 12.92, $p < .05$ (McNemar, 1962, p. 284) thus demonstrating that higher order unit sizes contribute i tantly to recall differences between g

Similar correlation comparisons also made between *S*s within each present 4 major groups, as shown i ble 2. Higher order units clearly acc for more of the within-groups vari only for the categorical list, particularly in the categorical–random condition, F (3, 25) = 11.95, $p < .001$. By contrast, higher order units contributed nothing within unrelated-list conditions.

Input–output. The major results of these analyses are shown in Figure 3. Organization increased over trials in all 4 unit-size analyses, *F*s (11, 1232) > 6.63. Only Unit Size 2 yielded an overall effect of list structure, F (1, 112) = 28.45. However, all 4 unit-size analyses produced a significant List Structure × Trials interaction, *F*s (11, 1232) > 4.59, reflecting the increasing superiority of unrelated conditions with practice. Figure 3 also shows large differences favoring fixed- over random-order conditions which were significant for all unit sizes, *F*s (1, 112) > 14.43. There was a significant List Order × Trials interaction in all analyses, *F*s (11, 1232)

TABLE 2

SIMPLE AND MULTIPLE CORRELATIONS BETWEEN MEAN RECALL AND ORGANIZATION UNIT-SIZE MEASURES

Condition	Unit size			
	2	2, 3	2, 3, 4	2, 3, 4, 5
Categoric–random	.45*	.45*	.70**	.82**
Categoric–fixed	.21	.25	.26	.34
Unrelated–random	.65**	.65**	.65**	.66**
Unrelated–fixed	.47*	.47*	.48*	.49*

* $p < .01$.
** $p < .001$.

> 12.48. As shown in Figure 3, this increasing fixed-random difference over trials was limited to unrelated-list conditions, yielding Fs (11, 1232) > 3.98.

The input–output data for the unrelated conditions are consistent with Mandler and Dean's (1969) results. However, the present level of input–output organization for categorical–fixed conditions is well above that reported by Mandler (1969), who concluded that no *S*s in this group showed a seriation tendency. The present data indicated that at least 2 of 30 *S*s in the categorical-fixed group showed perfect seriation, although this in itself could not account for the mean level of performance exhibited. Another difference from Mandler was the present failure to find any taxonomic clustering superiority of the categorical–fixed condition relative to the categorical–random condition which indicates that seriation may be used in lieu of or in addition to taxonomic organization in typical categorized lists. Mandler's failure to find a seriation tendency in this group is probably because his *S*s were completely instructed about the taxonomic list structure before learning.

Primacy–recency. The input–output analyses assess the role of serial input order as a general determinant of recall order, but provide no information about the relative position of these items in the recall order or recency and primacy determinants of recall order. To determine any such effects, an adaptation of the standarized recall rank measure (Battig et al., 1965) was used, whereby positive and negative values, respectively, indicate mean recall positions before or after the median on that trial. Although obtained for each fifth of the list on each trial, only the primacy and recency components showed significant effects, so these alone will be discussed here.

As shown in Figure 4, primacy and recency components showed no significant overall differences with respect to position in recall, but there was a significant Primacy–Recency × Trials interaction, F (11, 1232) = 11.03. All 4 panels of Figure 4 show this interaction to reflect

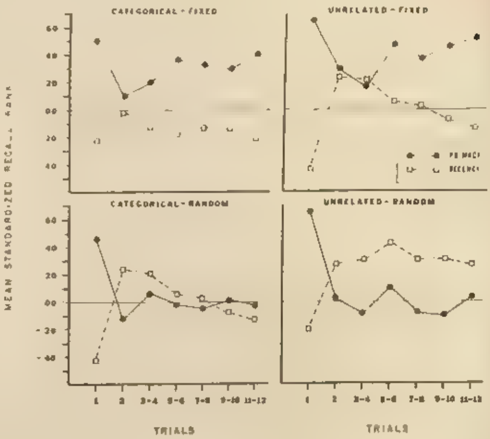


FIGURE 4. Mean standardized recall rank values for primacy and recency components of list order as a function of list structure, list order, and trials.

the large primacy–recency difference on Trial 1 which is reduced and reversed on Trial 2. Figure 4 additionally shows that this Primacy–Recency × Trials interaction depended upon list presentation order, F (11, 1232) = 2.52. Significant also was the opposite direction of this primacy–recency difference for fixed- (.43) and random-order conditions (−.15). The final significant effect was the overall difference between categorical and unrelated lists, F (1, 112) = 24.63, indicating that the pooled primacy and recency components occurred earlier in recall in the unrelated- (.18) than categorical-list (.07) conditions.

The results of this analysis are in general agreement with those studies (e.g., Deese & Kaufman, 1957) using correlational techniques to demonstrate a recency strategy. However, the present technique indicates that there are 2 factors determining the occurrence of such a strategy. The first factor is experience in the experimental situation, such that a naive S does not demonstrate any recency effect until Trial 2. The second factor concerns the maintenance of such a strategy which depends on both the list structure and list order, since consistent recency only occurs in unrelated–random conditions where no semantic or serial position cues are provided. The primacy component results

are in agreement with Mandler and Dean (1969) who found that unrelated constant serial order lists contain a constant primacy component, and also extend this result to categorical–constant serial order lists. The latter results also indicate the sensitivity of this organizational technique for studying primacy and recency effects.

Priority. The Battig et al. (1965) procedures were also applied to the analysis of priority of recall of newly learned items (PRNI). Items were considered as new items only if never recalled previously. This analysis was performed only on Trials 2–6 inclusive since no new items occurred following Trial 6. Included in this analysis were the variables of item type (old vs. new) and trials in addition to the experimental variables. Much higher priority was given to new items (.19) than to old items (−.19), F (1, 112) = 178.82. An Item Type × Trials interaction, F (4, 448) = 3.97, reflected increasing priority scores for new items across Trials 2–4 (.20, .25, and .34, respectively) which decreased on Trials 5 and 6 (.08 for both), while the priority scores for old items remained stable across all trials (—.22, —.20, —.18, —.17, —.18, respectively).[5] There was a List Order × Item Type interaction F (1, 112) = 10.06, reflecting no difference in the rank of old items for either type of list order, but a much higher priority for new items under the random- (.27) than fixed-order (.10) conditions. A similar List Structure × Item Type interaction, F (1, 112) = 6.79, resulted from a much higher priority given to new items under the unrelated (.26) than the categorized (.12) conditions but no such difference for old items.

The present PRNI effect suppo[rts] strongly several previous studies demo[n]strating such an effect (e.g., Battig [&] Slaybaugh, 1969). Of particular interest however, is the limitation of this PRNI effect both with categorical structure and fixed serial orders. As in the present analysis of recency, this mode of organization seems to suffer when semantic or serial list cues are present.

Discussion

The present analyses of higher order SO units have demonstrated that such units are formed in typical free-recall learning, and more importantly, that they vary in their sequential reproduction across trials. These data therefore support the contentions that pair-wise sequential indices are insensitive and underestimate total SO. The insensitivity problem was further illustrated by the 96%–45% difference in the amount of between-groups recall variance that could be accounted for by using all unit sizes as opposed to Size 2 alone.

The failure to observe similar correlation increases within groups, with the exception of the categorical–random condition, provides further information about the structure of recall and the relationships among unit-size measures. Since *S*s in fixed-order conditions relied on input order as a basis of organization, thereby reducing the internal sequential variability in organization units (see Table 1), any higher order organization should be almost completely measured by Size 2 units. Correlations between pair-wise and higher order unit sizes were above .925 thus supporting this interpretation. In the unrelated–random conditions it is probably the delayed development of higher order units rather than any strong relationships between unit-size measures (all rs $< .690$) that caused the correlations to remain constant. Support for this interpretation comes from similar analyses based on the last 2 trials whereby the correlation increased from .601 to .721 by the addition of Size 3 units, producing a marginally significant difference ($p < .10$). Stronger evidence that higher order units enhance between-*S*s organization-recall relationships within unrelated–random lists comes from a similar analysis of previous data (Rogers & Battig, 1972). This showed a correlation of −.001 for pair-wise units which increased significantly to .454 when units of Size 3 were included.

[5] In the standardized recall rank technique, sets of items comprising the total recall on that trial are compared with respect to position in recall. Only when the sets are nearly equal in size does the mean of the standardized recall rank values obtained across all sets necessarily approach 0. However, as the sets become increasingly disproportionate in size (as when old exceed new items after Trial 2), the mean standardized recall rank values across all sets can differ substantially from 0.

The overall correlational data indicate that fairly strong recall-organization relationships may operate for the higher order units, although substantial between-*S*s variability in recall within groups remains unaccounted for by such relationships. Part of this variability can be attributed to the existence of several different types of organizational processes which occur simultaneously, often in reciprocal relationship to one another, and which can affect recall. Evidence of such reciprocal relationships is provided by the findings of (*a*) more SO but less overall primacy, recency, and priority-of-new-item effects within categorical than unrelated lists, and (*b*) more SO for fixed- than random-order conditions but with less priority of new items and also a reduced recency strategy.

Besides showing an overall reciprocal relationship between processes, the data also indicate that such relationships change with practice. A hierarchy of organizational strategies or series of stages is evident such that *S*s begin with a strong seriation strategy on the first trial followed by recency and priority strategies, and culminating in higher order SO processes. The main difference between conditions appears to be in the second and third stages of this organizational development. In those conditions where there is an interitem semantic relationship, or a fixed serial order, or both, the degree to which priority and recency strategies are developed is limited and these *S*s appear to proceed more rapidly to an organizational stage involving the formation of higher order *S* units. Only in the unrelated–random condition do *S*s maintain a recency strategy and show little development of higher order SO units, thus contradicting Puff's (1972) conclusion that recency might facilitate SO in unrelated lists. It appears that higher order units per se are not sufficient to explain recall differences, and that overall speed or ease of transition to this stage is equally important. Without semantic or sequential cues, *S*s may be forced to operate with simpler organizational strategies.

Based on the foregoing evidence of the interactive nature of organizational processes and changes in such processes over trials, previous inconsistencies in the recall-organization literature appear due to at least 2 factors, i.e., (*a*) inability of any one index to account for the complete pattern of organization and (*b*) differences in this pattern depending on degree of practice.

An important additional consideration with respect to the present results concerns their relationship to storage and retrieval processes. One apparently unresolved issue is whether the organization observed at output is a function of storage, retrieval, or both. Slamecka's (1968, 1969) tests of the independent storage hypothesis were based upon list organization of items by pairs, i.e., the organizational units could be represented by direct associative relationships between pairs of list items. The present data question the validity of this assumption and his tests of the independence hypothesis. Additionally, given that higher order units are formed, one must question what constitutes the appropriate accessibility cue for any individual list item. The general lack of success of studies which have attempted to use individual list item cuing may be a function of this problem.

Given that higher order free-varying units are formed, one can also wonder which current theories of storage and retrieval would predict such structures. The development of such units is consistent with the hierarchical model of Mandler (1967) although he did not state how recall should occur within or between units. A random selection process at all levels of the hierarchy would produce the desired results and would be similar to Shiffrin's (1970) restricted search hypothesis. An alternative theory based upon an associative network (Anderson, 1972) might also produce organization results similar to those observed. In addition, Anderson's model would predict that the formation of higher order units is a direct result of storage and the elaboration of the associative graph. Since this graph changes from trial to trial, organizational units could be maintained although the order within such units might vary. The output protocols produced by Anderson's model would have to be analyzed for units larger than 2 to more fully determine the adequacy of the model.

Also, while it may be relatively easy to predict the sequential organization properties of recall protocols via random selection within and across units, it is more difficult to hypothesize how such random units are formed. Wallace (1970) has speculated that the pairwise organization typically measured is a function of implicit and explicit contiguity during learning. Such a theory must ultimately be able to explain the free-varying internal structure of higher order units in random-list-order conditions and the reduced variability under constant-order conditions. A comparison of Anderson's (1972) and

Wallace's associative theories seems to be required to help resolve how such units are formed, what circumstances dictate their formation, and whether any associative theory is capable of explaining them. Irrespective of specific issues, the present organization results certainly must be dealt with in any complete theory of storage and retrieval processes.

REFERENCES

ANDERSON, J. R. FRAN: A simulation model of free recall. In G. Bower (Ed.), *The psychology of learning and motivation.* Vol. 5. New York: Academic Press, 1972.

BATTIG, W. F., ALLEN, M., & JENSEN, A. R. Priority of free recall of newly learned items. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 175-179.

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph,* 1969, **80** (3, Pt. 2).

BATTIG, W. F., & SLAYBAUGH, G. D. Evidence [illegible]

[illegible] 1966, **19**, 935-942.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology,* 1953, **49**, 229-240.

[illegible] call of unorganized and sequentially organized [illegible] **54** [illegible]

[illegible]

MANDLER, G., & DEAN, P. J. Seriation: Development of serial order in free recall. *Journal of Experimental Psychology,* 1969, **81**, 207-215.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1962.

PELLEGRINO, J. W. A general measure of organization in free recall for variable unit size and internal sequential consistency. *Behavior Research Methods and Instrumentation,* 1971, **3**, 241-246.

PELLEGRINO, J. W. A FORTRAN IV program for analyzing higher-order subjective organization units in free recall learning. *Behavior Research Methods and Instrumentation,* 1972, 215-217.

POSTMAN, L. A pragmatic view of organization theory. In E. Tulving & W. Donaldson (Ed.), *Organization of memory.* New York: Academic Press, 1972.

PUFF, C. R. An investigation of two forms of organization in free recall. *Journal of Verbal Learning and Verbal Behavior,* 1970, **9**, 720-724.

PUFF, C. R. Temporal properties of organization in recall of unrelated words. *Journal of Experimental Psychology,* 1972, **92**, 225-231.

ROGERS, J. L., & BATTIG, W. F. Effect of amount of prior free recall learning on paired-associate transfer. *Journal of Experimental Psychology,* 1972, **92**, 373-377.

SHIFFRIN, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

SHUELL, T. J. Clustering and organization in free recall. *Psychological Bulletin,* 1969, **72**, 353-374.

[illegible] in free recall. *Journal of Experimental Psychology,* 1968, **76**, 504-513.

SLAMECKA, N. J. Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology,* 1969, **81**, 557-560.

TULVING, E. Subjective organization in free recall of "unrelated" words. *Psychological Review,* 1962, **69**, 344-354.

[illegible] free recall [illegible] Englewood Cliffs, N. J.: Prentice-Hall, 1968.

[illegible] *Verbal Behavior,* 1970, **9**, 58-68.

WINOGRAD, E., COHN, C. P., & RAND, J. Superiority of complete presentation to single-item presentation in recall of [illegible]. *Journal of Experimental Psychology,* 197? **88** [illegible]

(Received June 11, 1973)

Journal of Experimental Psychology [illegible]
[illegible] No. 3, 473-[illegible]

SEQUENTIAL EFFECTS IN DISCRETE-TRIALS INSTRUMENTAL ESCAPE CONDITIONING

JEFFREY A. SEYBERT, ROGER L. MELLGREN, JARED B. JOBE, AND ED ECKERT

University of Oklahoma

Five experiments investigated both acquisition and extinction effects in a discrete-trials (runway) escape paradigm. The escape procedure consisted of shocking *S* in the start and run sections of the runway, and reducing shock level in the goal box on reinforced (R) trials but not on nonreinforced (N) trials. The results of these experiments were: (*a*) greater resistance to extinction was exhibited by partially reinforced *Ss* than continuously reinforced *Ss*, (*b*) presence and number of N-R transitions influenced resistance to extinction, (*c*) number of successively nonreinforced trials (N length) was also a determinant of resistance to extinction, and (*d*) alternating N and R trials resulted in patterned running, but a double-alternation schedule did not. The results of all 5 experiments were interpreted as sup-[illegible] 1967 [illegible] extended to include instrumental escape conditioning phenomena.

Capaldi's (1967) sequential hypothesis of instrumental learning has received considerable support when appetitive instrumental conditioning procedures are used (e.g., Robbins, 1971). Relatively little work has been reported, however, investigating sequential variables in aversive conditioning paradigms, although there is some evidence to indicate that sequential factors are important in determining response persistence in punishment situations (e.g., Campbell, Wroten, & Cleveland, in press; Capaldi & Levy, 1972; Dyck, Mellgren, & Nation, 1974). It was the purpose of the present experiments to investigate the operation of sequential variables in instrumental escape conditioning and to determine if it is possible to extend the boundary conditions of sequential theory to an area of research where they previously have not been applied.

Recently, Franchina and his co-workers (e.g., Franchina & Snyder, 1970) have reported several escape conditioning experiments in which the sequence of shock and nonshock trials was manipulated in a hurdle-box apparatus. A shock trial consisted of shock present in 1 side of the apparatus, and a nonshock trial consisted of the absence of shock in both sides of the apparatus. These authors and others (e.g., Woods, Markman, Lynch, & Stokely, 1972) have noted, however, that these data cannot be explained in terms of the sequential effects of reinforcement and nonreinforcement because the shock-nonshock procedure is not analogous to reward-nonreward trials in the appetitive situation, since there is a lack of primary motivation on the nonshock trials. Thus, in order to test the effects of sequential variables in instrumental escape conditioning, it would be necessary to utilize procedures analogous to rewarded and nonrewarded trials in appetitive instrumental conditioning. For example, in a runway, a "rewarded" escape trial would consist of the application of some aversive stimulus to *S* in the start and run sections of the straight runway and the absence (or reduced level) of that aversive stimulus in the goal box. A "nonrewarded" escape trial would consist of the application of the same aversive stimulus in all sections of the runway. It should be noted that the duration of nonreward must necessarily be limited in this procedure because at some time after *S* enters the goal box it must be removed. Thus, the above procedure may be more analogous to delay of reward, but since both nonreward and delay have been postulated to involve similar theoretical mechanisms (Capaldi,

[1] Requests for reprints should be sent to Roger L. Mellgren, Department of Psychology, University of Oklahoma, Norman, Oklahoma [illegible]

1967) the terminology of nonreward will be maintained.

One prediction of sequential theory that has been confirmed in appetitive studies involves the Schedule of Reward × Number of Acquisition Trials interaction. Following limited acquisition, reward schedules with N lengths—number of consecutive nonreinforced (N) trials followed by a reinforced (R) trial—of 1 produce greater resistance to extinction (Rn) than do schedules with longer N lengths; but after extended acquisition training, reward schedules with long N lengths produce greater Rn than do schedules with N lengths of 1 (see Capaldi, 1967) when percentage of reinforcement is equated for the 2 groups. Experiments I and II represent an attempt to demonstrate this interaction using an escape conditioning procedure.

Experiment I

Following limited acquisition training, a group receiving N lengths of 1 should show greater Rn than a group receiving N lengths of 3, because there should be greater strength of the association between the memory of nonreward and the instrumental response (S^N–R_I in sequential terminology) for the N length 1 group due to the greater number of transitions from nonreinforcement to reinforcement (N–R transitions) that occur when percentage of reinforcement is equated. Both partial reinforcement (PRF) groups should show more Rn than the continuous reinforcement (CRF) control group.

Method

Subjects. The *S*s were 36 experimentally naive male albino rats of the Sprague–Dawley strain, purchased from the Holtzman Co. They were approximately 100 days old at the start of training and were randomly assigned to 1 of 3 groups (n = 12/group).

Apparatus. The apparatus consisted of a straight-alley runway manufactured by the Hunter Co. The alley was constructed of clear Plexiglas with a grid floor and was 150 × 15 × 9 cm. It was divided into a 30-cm. start section, a 90-cm. run section, and a 30-cm. goal section, with all sections separated by guillotine doors. The *S*s' progress in the alley was measured by 3 .01-sec. Standard timers: the first, which measured start time, was started by a microswitch at the start-box door and stopped by a photocell located 11 cm. into the alley; the second, which measured run time, was started by the first photocell and stopped by a second photocell located 11 cm. in front of the goal box; the third, which measured goal time, was started by the second photocell and stopped by a third photocell located 9 cm. inside the goal box. Start, run, and goal speeds were obtained by converting the start, run, and goal times to reciprocals. A .3-ma. scrambled shock was automatically delivered to the start and run sections of the alley by a Model 700 Grason Stadler shock generator, which was activated by the microswitch at the start-box door. Shock of the same strength could be administered manually to the goal section of the alley by means of a switch mounted on the apparatus control panel. This apparatus was also used in Experiment II.

Procedure. Upon arrival in the laboratory, *S*s were individually housed and placed on an ad-lib schedule of food and water which continued throughout the experiment. During the 3 days before the start of the experiment, each *S* was handled for 5–10 min. daily. One day of pretraining preceded the actual start of the experiment. During this period, all *S*s received 2 R trials in the apparatus. The following procedure was employed throughou[illegible] the experiment on R trials. The *S* was pl[illegible] in the start box, and after a peri[illegible] start-box door was raised, activat[illegible] timer and the shock generator. [illegible] traversed the runway and entere[illegible] trified goal box, the goal-box door was closed and *S* was confined for 15 sec., after which it was removed and placed in a carrying cage. The procedure on N trials was simila[illegible] R trials except that as soon as *S* left the [illegible] box, *E* depressed the switch which electrified the grid floor of the goal box with a .3-ma. shock. When *S* entered the goal box, the door was closed and it was confined to the electrified goal box for 15 sec., after which it was removed to the carrying cage. Thus, reinforcement consisted of a 15-sec. confinement in the nonelectrified goal box following shock in the start and run sections, while nonreinforcement consisted of a 15-sec. confinement in the electrified goal box following shock in the start and run sections.

Acquisition training consisted of 8 trials per day for 2 days, a total of 16 acquisition trials. Two groups received a 62% PRF schedule in the escape apparatus, while the third group (Group CRF) was reinforced on all trials during acquisition. Group 1N had 3 N trials in each daily session, each N trial being followed by at least 1 R trial. The acquisition reward schedule for Group 1N was as follows: Day 1, RRNRNRNR; and Day 2, RNRNRRNR. Group 3N also received 3 N trials in each daily session, but the N trials occurred consecutively. The acquisition reward schedule for Group 3N was as follows: Day 1, RRNNNRRR; and Day 2, RRRNNNRR. The *S*s were run in

squads of 3 with an intertrial interval (ITI) of 3–4 min.

After acquisition training, all *S*s were given 4 extinction trials per day for 3 days, a total of 12 extinction trials. The procedure for extinction trials was the same as for the N trials during acquisition.

Results and Discussion

In the 5 experiments reported here, data from all runway sections were examined. However, the most consistent and reliable effects were observed in the start-speed measure, and therefore all analyses reported and discussed in this and subsequent experiments will be in reference to start speeds.

An analysis performed on the last block of 8 acquisition trials indicated no terminal acquisition differences, $F\ (2,\ 33) < 1$. A 3×4 (Groups $\times$ Days) analysis of variance performed on the start-speed data for the last day of acquisition and the 3 days of extinction revealed a significant days effect, $F\ (3,\ 99) = 3.96$, $p < .025$; a marginally significant groups effect, $F\ (2,\ 33) = 2.82$, $p = .07$; and a nonsignificant Groups $\times$ Days interaction, $F\ (6,\ 99) < 1$. Sequential theory makes a specific prediction concerning the rank ordering of the groups following limited acquisition: Group 1N should demonstrate greater Rn than Group 3N, and Group 3N should be more resistant than Group CRF. This prediction was tested by a Jonckheere (1954) test. The Jonckheere procedure tests the null hypothesis that groups are not significantly ranked in a specific predicted order, and thus it is ideally suited for examining the predictions of sequential theory in the present experiment. Results of the Jonckheere test performed on the sums of start speeds for the 3 days of extinction indicated than the groups demonstrated the predicted rank ordering, $z = 2.20$, $p = .01$. The data from Experiment I may be seen in Figure 1.

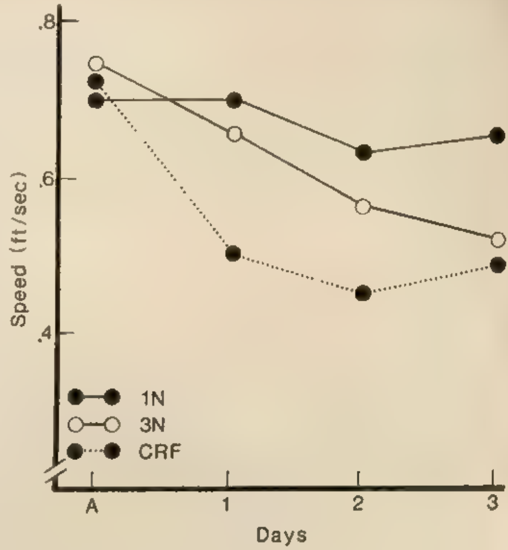


FIGURE 1. Mean start speeds for the last day of acquisition (A) and 3 days of extinction.

The results of Experiment I revealed that one portion of the Schedule of Reward $\times$ Number of Acquisition Trials interaction found in appetitive conditioning also occurs in instrumental escape conditioning. Specifically, following a limited number of acquisition trials, a group receiving short N lengths in acquisition was more resistant to extinction than a group receiving longer N lengths during acquisition. This finding supports the prediction made from sequential theory.

EXPERIMENT II

According to sequential theory, after extended acquisition, it would be predicted that a group receiving N lengths of 3 in the instrumental escape procedure should show more Rn than a group receiving N lengths of 1, because the strength of the hypothetical S^N–R_I association would be asymptotic for both groups, and the N Length 3 group would experience less generalization decrement in extinction.

Method

Subjects. The *S*s were 36 experimentally naive male albino rats of the Sprague-Dawley strain purchased from the Holtzman Co. They were approximately 85 days old at the start of the experiment and were randomly assigned to 1 of 3 groups ($n = 12$/group).

Procedure. After arrival in the laboratory, *S*s were individually housed and placed on ad-lib food and water. As in Experiment I, *S*s were handled individually for 3 days prior to the start of the experiment and were given 2 reinforced trials in

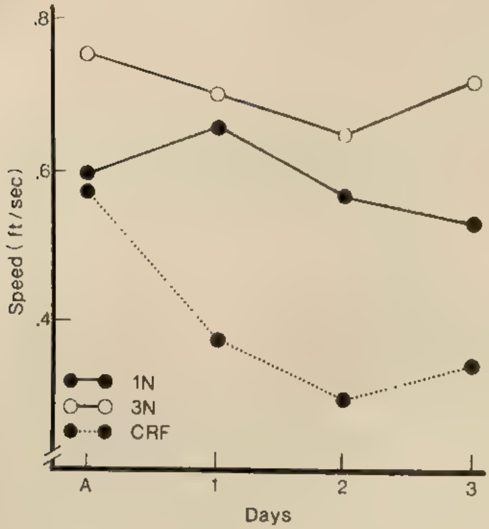


FIGURE 2. Mean start speeds for the last day of acquisition (A) and 3 days of extinction.

the runway as pretraining. The procedure for R and N trials was exactly the same as in Experiment I.

Acquisition training consisted of 8 trials per day for 5 days, a total of 40 acquisition trials. Two groups had a 62% PRF schedule, while the third group (Group CRF) was reinforced on all trials in acquisition. Group 1N had 3 N trials in each daily session, each N trial being followed by an R trial (3 N lengths of 1). Group 3N also had 3 N trials in each daily session, but the N trials occurred consecutively (1 N length of 3). The actual sequence of trials in acquisition for all groups may be seen in Table 1. Throughout the experiment, *Ss* were run in squads of 3 (1 from each group), with the ITI being approximately 3–4 min. for each *S*.

Upon completion of acquisition training, *Ss* were given 12 extinction trials, 4 per day for 3 days. The procedure for all extinction trials was the same as for the N trials during acquisition.

Results and Discussion

The analysis performed on start speeds for the last block of 8 trials in acquisition

TABLE 1

SEQUENCE OF REINFORCED (R) AND NONREINFORCED (N) TRIALS FOR ALL GROUPS DURING ACQUISITION

Day	Group		
	1N	3N	CRF
1,5	RRNRNRNR	RRNNNRRR	RRRRRRRR
2	RNRNRRNR	RRRNNNRR	RRRRRRRR
3	RNRRNRNR	RNNNRRRR	RRRRRRRR
4	RNRNRNRR	RRRRNNNR	RRRRRRRR

indicated no terminal acquisition differences, F (2, 33) < 1. A 3 × 4 (Groups × Days) analysis of variance performed on the start speeds for the last day of acquisition and the 3 days of extinction revealed a significant main effect for days, F (3, 99) = 3.51, p < .02, as well as a significant groups main effect, F (2, 33) = 3.89, p < .03. The Groups × Trials interaction failed to reach an acceptable level of significance, F (6, 99) = 1.38, p > .10. The Jonckheere test performed on the sums of the start speeds for 3 days of extinction indicated that the groups were rank ordered as predicted, z = 3.00, p = .001. The data for Experiment II may be seen in Figure 2.

The results of Experiment II indicated that the second part of the Schedule of Reward × Number of Acquisition Trials interaction, commonly found in appetitive conditioning and predicted by sequential theory, is also observed in instrumental escape conditioning. Following extended acquisition, the group receiving N lengths of 3 showed greater Rn than the group receiving N lengths of 1.

EXPERIMENT III

Although the results of Experiments I and II were statistically reliable, they might be seen as somewhat unimpressive, and it was deemed desirable that these effects be systematically replicated and extended, as there is a paucity of data concerning extinction effects in escape conditioning. Therefore, Experiment III represents an attempt to again demonstrate the effects of differential N lengths following limited acquisition, using different procedural parameters.

The procedural changes in the present experiment were made on the basis of unpublished work in the authors' laboratory which indicated that such changes might produce more clearly delineated results. These changes were: (*a*) an increase in the level of shock, (*b*) an increase in the reinforcing amount of shock reduction, (*c*) the presence of a very low level of shock in the goal box on R trials, and (*d*) an increase in the goal-box confinement time

on both R and N trials. Finally, the continuously reinforced control group was omitted, since the PRF extinction effect in escape conditioning has been reliably demonstrated, both in Experiments I and II reported here and elsewhere (e.g., Bower, 1960; Woods et al., 1972).

Method

Subjects. The *Ss* were 16 experimentally naive male albino rats of the Sprague-Dawley strain, purchased from the Holtzman Co. They were approximately 110 days old at the start of the experiment and were randomly assigned to 1 of 2 groups (n = 8/group).

Apparatus. The apparatus was identical to that used in Experiment I with 1 exception. An automatic timer was included in the control panel to regulate shock and confinement time in the goal box. This apparatus was also used in Experiments IV and V.

Procedure. Upon arrival in the laboratory, *Ss* were housed 2 to a cage and placed on an ad-lib schedule of food and water which continued throughout the experiment. One day prior to the start of the experiment, each *S* was given 1 reinforced pretraining trial in the apparatus. The procedure on R trials was similar to that in the first 2 experiments except that the shock level in the start and run sections was .5 ma., the shock level in the goal box was .1 ma., and the goal-box confinement time was 30 sec. On N trials the shock level was the same (.5 ma.) in all sections, and goal-box confinement time was 30 sec. Thus, reinforcement consisted of a 30-sec. confinement in the .1-ma. electrified goal box following .5-ma. shock in the start and run sections (a .4-ma. shock reduction), while nonreinforcement consisted of a 30-sec. confinement in the .5-ma. electrified goal box following .5 ma. in the start and run sections (no shock reduction).

As in Experiment I, acquisition training consisted of 8 trials per day for 2 days, a total of 16 trials. The 2 groups (Group 1N and Group 3N) received 62% PRF in the apparatus, with the same schedules used for Group 1N and Group 3N as in Experiment I. The *Ss* were run in squads of 4 (2 *Ss* from each group per squad), with the ITI for each *S* being approximately 4 min. Following acquisition training, all *Ss* received 8 extinction trials per day for 2 days, a total of 16 extinction trials. The procedure for extinction trials was exactly the same as that for N trials during acquisition.

Results and Discussion

An analysis performed on start speeds for the last block of 8 acquisition trials indicated that there were no terminal acquisition differences, t (14) = .70, p > .05.

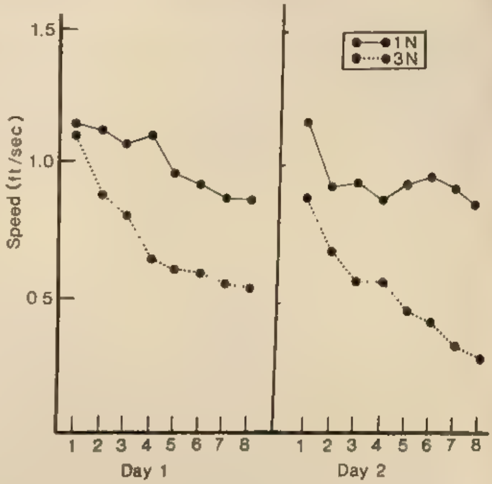


FIGURE 3. Mean start speeds for 2 days of extinction (8 trials/day, plotted trial by trial).

It can be seen from Figure 3 that the extinction results of this experiment closely parallel those of Experiment I and are also comparable to appetitive studies using similar procedures (see Capaldi, 1967): Group 1N showed greater response persistence than did Group 3N on both days of extinction. These conclusions were confirmed by separate 2 × 8 (Groups × Trials) analyses of variance performed on the start-speed data for each day of extinction. On Day 1 the groups main effect was significant, F (1, 14) = 5.24, p < .05; as well as the trials main effect, F (7, 98) = 8.44, p < .001; but the Groups × Trials interaction failed to reach an acceptable level of significance, F (7, 98) = 1.74, p > .10. On Day 2 the groups main effect was again significant, F (1, 14) = 15.74, p < .005, as was the trials main effect, F (7, 98) = 17.50, p < .001. Since there was a large difference between the groups on Trial 1 of Day 2, a significant Groups × Trials interaction would be necessary in order to demonstrate differential *rates* of extinction as opposed to simply differential amounts of spontaneous recovery. This interaction was significant, F (7, 98) = 7.92, p < .001.

Thus the finding that, following limited acquisition, short N lengths produce greater Rn than long N lengths is reliable

in instrumental escape conditioning. It should also be noted that the changes in procedure from Experiment I to the present experiment, i.e., increasing shock level, amount of shock reduction, and goal-box confinement time, seem to increase the magnitude of these effects.

Experiment IV

It has been demonstrated by a number of investigators using appetitive procedures (see Capaldi, 1967) that rats, if provided with a regular sequence of R and N trials, will learn to respond appropriately. That is, they will learn to run slowly on N trials and fast on R trials. It was the purpose of this experiment to investigate this phenomenon, known as patterned running or patterning, in an instrumental escape conditioning paradigm. Patterning has been demonstrated in appetitive conditioning when single-alternation schedules are used, i.e., alternation of N and R trials; however, patterning is not usually observed when double-alternation schedules are used, i.e., 2 N trials alternated with 2 R trials (e.g., Bloom & Capaldi, 1961). The present experiment utilized both single- and double-alternation schedules in order to determine if similar results hold using the instrumental escape procedure.

Method

Subjects. The *Ss* were 35 male albino rats of the Sprague–Dawley strain, obtained from the Holtzman Co. They were approximately 180 days old at the start of the experiment and were randomly assigned to 1 of 5 groups ($n = 7$/group).

Procedure. Throughout the experiment *Ss* were housed individually and maintained on an ad-lib schedule of food and water. One day prior to the start of the experiment, all *Ss* were given 2 reinforced pretraining trials in the apparatus. Reinforced and nonreinforced trials were conducted exactly as in Experiment III.

Acquisition training consisted of 4 trials per day for 30 days, a total of 120 acquisition trials. Four groups received a 50% PRF schedule during acquisition, while a fifth group (Group CRF) was reinforced on all trials during acquisition. The 4 PRF groups formed the cells of a 2 × 2 factorial with the factors being N length (1 or 2) and type of ordinal trial (N trial first or R trial first). Thus, the 4 PRF groups were labeled NRNR, RNRN, NNRR, or RRNN to denote their schedule of R and N trials on each day of acquis[...] were run in squads of 5, 1 *S* from [...] a squad, with an ITI of 5–6 min. for [...]

Following acquisition training, all [...] 4 extinction trials per day for 10 da[...] 40 extinction trials. The procedure [...] trials was exactly the same as that [...] trials during acquisition.

Results and Discussion

Acquisition. An analysis on the mean of all 4 trials of the last day in acquisition indicated no differences between the groups at the end of acquisition, F (4, 30) < 1. The start-speed data from the PRF groups for all days of acquisition were analyzed in order to determine if any patterning occurred. So that the data would not be contaminated by any possible first-trial warmup effect, the first trial of each day was not included in the analysis. The fourth trial of each day was also deleted in order to equate the number of N and R trials in the analysis. Thus, only the 2 middle trials for each day were analyzed. These data were combined into 3-day blocks and analyzed via a 2 × 2 × 2 × 10 (N Length × Type of Ordinal Trial—N First or R First × Type of Trial—N or R × Trial Blocks) analysis of variance. The type of ordinal trial main effect was nonsignificant, F (1, 24) < 1, and it did not interact with any other variables. The N length main effect also failed to reach an acceptable level of significance, F (1, 24) < 1, although the N Length × Type of Trial interaction was significant, F (1, 24) = 4.17, $p < .05$, as well as the N Length × Type of Trial × Blocks interaction, F (9, 216) = 3.52, $p < .001$. The main effect of type of trial was significant, F (1, 24) = 15.03, p = .001, as was the blocks main effect, F (9, 216) = 10.30, $p < .001$. Post hoc analysis (using the Scheffé correction procedure) of the 3-way interaction indicated patterning in the single-alternation groups (R trials faster than N trials) on Blocks [...] and 9, ps < .05. The *Ss* in the double-alternation groups did not differ in speed on N and R trials except on Block [...] (R trials faster than N trials, $p < .0$[...] Thus, in the present experiment the [...]

dence for patterning must be considered marginal, since faster running on R trials for the single-alternation groups was observed on only 2 trial blocks. There seem to be 3 possible explanations for the failure to find reliable patterning behavior: (*a*) It is possible that an insufficient number of acquisition training trials were run, since the patterning that was observed occurred late in acquisition. (*b*) The ITI during acquisition may have been too long (5–6 min.) for patterning to occur. In the appetitive situation, some investigators have observed patterning at long ITIs (e.g., Capaldi & Lynch, 1966), but others have failed to do so (e.g., Surridge & Amsel, 1966). Although a 5–6-min. ITI cannot be considered extremely long, it is nonetheless, substantially longer than the 15–30-sec. ITIs that are often used in sequential experiments. (*c*) Acquisition training may have consisted of too few trials in each daily session (4 per day) for patterning to develop. Bloom and Capaldi (1961), for example, demonstrated patterning by single-, but not double-alternation *S*s; they used a procedure that employed 12 trials per day, considerably more than in the present experiment. Thus, it is possible that any of these factors, or a combination of them, were responsible for the lack of patterning in the present data; or that patterning is difficult to obtain in escape conditioning.

Extinction. The start-speed data from extinction were blocked in 4-trial blocks (days) and analyzed via a 5 × 10 (Group × Days) analysis of variance. The extinction data may be seen in Figure 4. The analysis revealed significant main effects for groups, F (4, 30) = 60.58, $p < .001$, and days, F (9, 270) = 52.24, $p < .001$, as well as a significant Groups × Days interaction, F (36, 270) = 15.02, $p < .001$. Post hoc analysis (using the Scheffé procedure) indicated that all PRF groups were more resistant to extinction than Group CRF on all days (all ps $< .01$). There was no difference in the 2 single-alternation groups (NRNR and RNRN) throughout extinction. Group NNRR was more resistant to extinction than all other

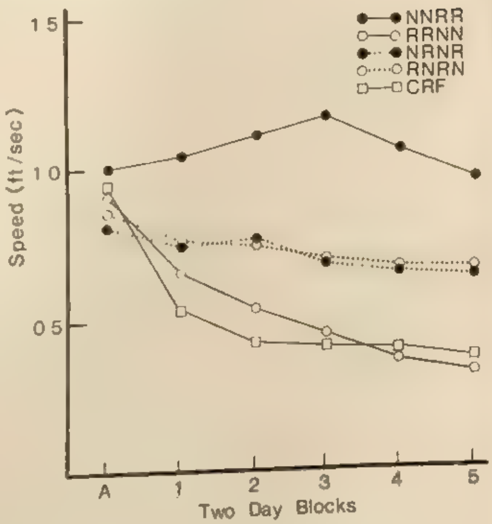


FIGURE 4. Mean start speeds for the last day of acquisition (A) and 10 days of extinction (in 2-day blocks).

groups for all days of extinction (all ps $< .01$). The RRNN group was inferior to Groups NRNR and RNRN on all days of extinction (all ps $< .01$) except Day 3, where no differences in the 3 groups were observed. Groups RRNN and CRF did not differ in Rn except on Days 1 and 3, where RRNN was faster than CRF ($p < .01$ in both cases).

The extinction results of the present experiment are consistent with those of Experiment II, i.e., following extended acquisition, a long-N-length group (Group NNRR) was more resistant to extinction than a short-N-length group (Groups NRNR and RNRN). The present experiment also graphically illustrates the importance of another sequential variable, N–R transitions, in determining response persistence. Group RRNN received an N length of 2 in the present experiment, but received no N–R transition in its daily trial schedule (compared with Group NNRR, which received an N length of 2 plus an N–R transition on each day of acquisition) and showed only slightly more Rn than Group CRF. This comparison demonstrates that in order for N length to be a factor in increased Rn, it must be followed by at least 1 R trial. This finding

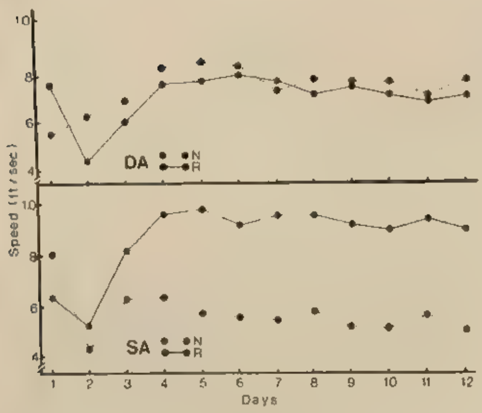


FIGURE 5. Mean start speeds on reinforced (R) and nonreinforced (N) trials for 12 days of acquisition.

is similar to results from appetitive experiments using similar reward schedules (Spivey & Hess, 1968; Spivey, Hess, & Black, 1968).

EXPERIMENT V

This experiment was designed to further investigate patterning behavior in instrumental escape conditioning. In an attempt to provide more favorable conditions for patterning to occur, a number of changes were made in the procedure. Specifically, more trials were run per day for each *S* (8 instead of 4), and the ITI was shortened considerably (45 sec. rather than 5–6 min.).

Method

Subjects. The *S*s were 24 male albino rats of the Sprague-Dawley strain, purchased from the Holtzman Co. They were approximately 125 days old at the start of the experiment and were randomly assigned to 1 of 3 groups (n = 8/group).

Procedure. Throughout the experiment, *S*s were housed individually and maintained on an ad-lib schedule of food and water. One day prior to the start of the experiment, all *S*s were given 2 reinforced pretraining trials in the apparatus. Reinforced and nonreinforced trials were conducted exactly as in Experiment III.

Acquisition training consisted of 8 trials per day for 12 days, a total of 96 acquisition trials. Two groups received a 50% PRF schedule during acquisition, while a third group (Group CRF) was reinforced on all trials during acquisition. The PRF groups differed only in the schedule of N and R trials received during acquisition. The single-alternation (SA) group received an NRNRNR R trial schedule on all 12 days of acquisition, while the double alternation (DA) group received an NNRRNNRR trial schedule on all 12 days. The *S*s were run in squads of 6 (2 *S*s from each group in each squad) with an ITI of 45 sec. for each *S*.

Following acquisition training, all *S*s were given 8 extinction trials per day for 5 days, a total of 40 extinction trials. The procedure on extinction trials was exactly the same as that for N trials during acquisition.

Results and Discussion

Acquisition. An analysis of variance performed on the mean of all 8 trials for the last day of acquisition indicated no terminal acquisition differences, F (2, 21) < 1. In order to examine possible patterning behavior the start-speed data from the 2 PRF groups were combined into 8-trial blocks (days) and analyzed by a 2 × 2 × 12 (Type of Trial—N or R × N Length × Days) analysis of variance. The main effect of type of trial was not significant, F (1, 14) < 1, although the Type of Trial × N Length interaction, F (1, 14) = 81.00, p < .001, and the of Trial × N Length × Days interaction, F (11, 154) = 6.71, p < .001, were significant. The N length main effect was significant, F (1, 14) = 37.12, p < .001, as was the days main effect, F (11, 154) = 3.82, p < .001. The N Length × Days interaction was significant, F (11, 154) = 2.43, p < .01, while the Type of Trial × Days interaction failed to reach an acceptable level of significance, F (11, 154) < 1. Post hoc analysis (using the Scheffé procedure) indicated that patterning began to appear on Day 3 (R trials faster than N trials, p < .05) and was reliable thereafter (Days 4–12, ps < .01) for Group SA. The only differences observed in Group DA were on Day 1 (R trials faster than N trials, p < .01) and Day 2 (N trials faster than R trials, p < .01); thereafter, there were no statistically reliable differences in speeds on N and R trials for Group DA. The acquisition data from Experiment V may be seen in Figure 5.

Thus, the present results demonstrate that reliable patterning behavior can occur in instrumental escape conditioning, although it seems to appear much earlier in

training than in similar appetitive procedures, and that massing of trials and/or fairly short ITIs are sufficient conditions for its occurrence.

Extinction. A 3 × 5 (Groups × Days) analysis of variance performed on start speeds revealed a significant groups main effect, F (2, 21) = 58.94, $p < .001$, and a significant days main effect, F (4, 84) = 2.71, $p < .05$, while the Groups × Days interaction failed to reach an acceptable level of significance, F (8, 84) < 1. Post hoc analysis (again using the Scheffé correction procedure) indicated that the groups were ordered in terms of Rn as follows: DA > SA > CRF (all ps < .01). The extinction data for Experiment V may be seen in Figure 6. The extinction results of the present experiment are consistent with those of Experiments II and IV in demonstrating that, following extended acquisition training, *S*s receiving longer N lengths are more resistant to extinction than *S*s receiving shorter N lengths. Even though *S*s in Group SA patterned, they were more resistant to extinction than Group CRF; this result is similar to extinction data in a number of appetitive patterning studies (e.g., Rudy, 1971).

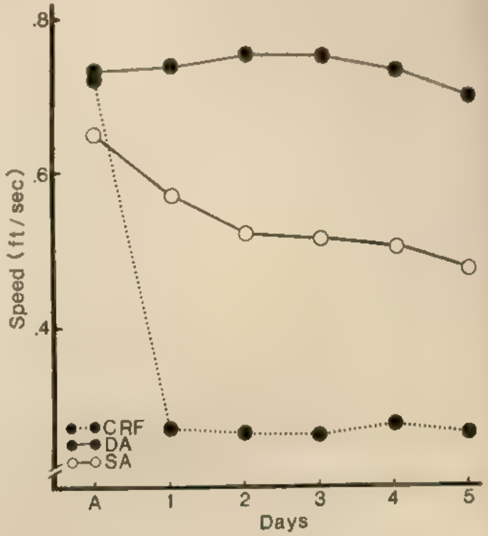


FIGURE 6. Mean start speeds for the last day of acquisition (A) and 5 days of extinction.

GENERAL DISCUSSION

Two general conclusions may be drawn from the data presented herein: (*a*) A reliable partial reinforcement extinction effect (PRE) was observed in all 4 experiments where a CRF control group was employed (Experiments I, II, IV, and V). The PRE is, of course, also reliably observed in appetitive instrumental conditioning, and has been previously demonstrated in escape conditioning (e.g., Bower, 1960; Woods et al., 1972). (*b*) In all 4 experiments where a CRF control group was used (Experiments I, II, IV, and V), there were no differences between the PRF and CRF groups at the end of acquisition. This result is in disagreement with some studies (e.g., Bower, 1960; Woods et al., 1972, with large magnitude of reinforcement) which have shown superiority for a CRF group at the end of acquisition. The reasons for this discrepancy are not clear at the present time. The present data are also discrepant in this respect with much of the appetitive conditioning literature in that PRF *S*s are often superior to CRF *S*s at the end of acquisition.

Several statements can also be made in relation to the effects of specific sequential variables in instrumental escape conditioning.

1. All 5 of the present experiments demonstrate N length effects very similar to those found in sequential appetitive conditioning studies. After limited acquisition, *S*s receiving short N lengths show more Rn than *S*s receiving long N lengths (Experiments I and III); and the opposite is true following extended acquisition (Experiments II, IV, and V). Thus the Schedule of Reward × Number of Acquisition Trials interaction seems to be a reliable finding in instrumental escape conditioning.

2. The occurrence of N–R transitions was shown to be important in escape conditioning, much as it has been in appetitive procedures. The effects of N–R transitions were most pronounced in Experiment IV where 1 double-alternation group (Group NNRR), which received 1 N–R transition per day, demontrated the greatest resistance to extinction, while the second double-alternation group (RRNN), which received no N–R transitions, was only slightly more resistant to extinction than the CRF control group. The slight superiority of Group RRNN over Group CRF can be accounted for by the fact that on each day of acquisition *S*s in Group RRNN

received an N–R transition with an ITI of 24 hr. Mellgren and Seybert (1973) have shown that N–R transitions with a long ITI in acquisition can increase massed-trials extinction performance. According to sequential theory, the number of N–R transitions is also postulated as the theoretical determinant of Rn after limited acquisition in appetitive situations (the number of N–R transitions determines the habit strength of the S^N–R_I association that determines Rn) and since the empirical data resulting from appetitive procedures and the present escape experiments are so similar, there is no reason to assume any different mechanism in escape conditioning. Thus, the results of Experiments I and III also implicate the importance of N–R transitions in determining Rn in escape conditioning. A more complete theoretical account of the sequential mechanisms that determine Rn after both limited and extended acquisition may be found elsewhere (e.g., Capaldi, 1967).

3. Pattern of N and R trials was shown to be an important variable in escape conditioning in Experiment V, where *S*s which received a regular single alternation schedule (Group SA) learned to run fast on R trials and slowly on N trials. These data are similar to those obtained when single-alternation schedules are used in appetitive in- [illegible] ing occurred earlier in escape training (after approximately 30 trials) than is usual with appetitive procedures.

Thus, the results of the present experiments demonstrate the operation of a number of [illegible] conditioning. These sequential variables were found to operate in escape conditioning in much the same way as they do in corre- [illegible]ng appetitive procedures. Therefore, the results may be seen as support for the hypothesis that Capaldi's (1967) sequential [illegible] mental escape conditioning phenomena. The present experiments join several others (e.g., Campbell et al., 1973; Capaldi & Levy, 1972; Dyck et al., 1974) in demonstrating that the boundary conditions of sequential theory may be expanded to include a number of aversive conditioning situations.

It should be noted, however, that some differences do exist between the present data [illegible] experiments. In appetitive experiments, the effects of sequential manipulations are usually observed in all runway sections (i.e., start, run, and goal). In the present escape conditioning experiments, the differences between groups as a result of the manipulation of sequential variables occurred most reliably in the start section. Although in some of the experiments differences did occur in other runway sections (e.g., the goal section in Experiment I, the run section in Experiment II, and the run and goal sections in Experiment III), these effects were most apparent in the start-speed measure. In Experiments IV and V, however, differences were observed in *only* the start section. Capaldi and Levy (1972), using punishment superimposed on a partial food reward schedule, also observed sequential effects predominantly in the start section of a straight runway. They attributed their results to a heightened avoidance gradient resulting from the highly aversive goal situation. The present data may be due to similar factors, although the punishment and escape procedures are somewhat different, particularly in relation to the locus and duration of the shock received by *S*. A second possible explanation for the present find[illegible] may lie in the number of acquisition tra[illegible] trials (120 in Experiment IV and 96 in [illegible]periment V). After large numbers of trials, the running response, as it occurs in the presence of shock, may become rather insensitive to other variables. Therefore, after relatively extended acquisition training, the main effect of acquisition variables may be restricted to the likelihood of the initiation of the running response, rather than speed of running after the response is initiated. This difference may be of considerable theoretical [illegible] and should be investigated [illegible]

REFERENCES

Bloom, J. M., & Capaldi, E. J. The behavior [illegible] patterns of partial reinforcement. *Journal of Comparative and Physiological Psychology*, 1961, 54, 261–265.

B[illegible] C. H. [illegible] related reward in escape learning. *Journal of Experimental P[illegible]*, [illegible] 59, [illegible]

Campbell, P. E., Wroten, J. D., & Cleveland, C. T. Re[illegible] following inter-mittent punishment and partial reinforcement: Effects of trial sequence. *Learning and Motivation*, in [illegible]

Capaldi, E. J. A sequential hypothesis of [illegible] In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 1. New York: Academic Press, 1967.

CAPALDI, E. J., & LEVY, K. J. Stimulus control of punished reactions: Sequence of punishment trials and magnitude of reinforcement trials. *Learning and Motivation*, 1972, **3**, 1–19.

CAPALDI, E. J., & LYNCH, D. Patterning at 24-hour ITI: Resolution of a discrepancy more apparent than real. *Psychonomic Science*, 1966, **6**, 229–230.

DYCK, D. G., MELLGREN, R. L., & NATION, J. R. Punishment of appetitively reinforced instrumental behavior: Factors affecting response persistence. *Journal of Experimental Psychology*, 1974, **102**, 125–132.

FRANCHINA, J. J., & SNYDER, C. R. Effect of patterns of shock and nonshock training trials on response alternation and extinction in escape training. *Psychonomic Science*, 1970, **21**, 177–179.

JONCKHEERE, A. R. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

MELLGREN, R. L., & SEYBERT, J. A. Resistance to extinction at spaced trials using the within-subject procedure. *Journal of Experimental Psychology*, 1973, **100**, 151–157.

ROBBINS, D. Partial reinforcement: A selective review of the alleyway literature since 1960. *Psychological Bulletin*, 1971, **76**, 415–431.

RUDY, J. W. Sequential variables as determiners of the rat's discrimination of reinforcement events: Effects on extinction performance. *Journal of Comparative and Physiological Psychology*, 1971, **77**, 476–481.

SPIVEY, J. E., & HESS, D. T. Effect of partial reinforcement trial sequences on extinction performance. *Psychonomic Science*, 1968, **10**, 375–376.

SPIVEY, J. E., HESS, D. T., & BLACK, D. Influence of partial reinforcement pattern and intertrial reinforcement on extinction performance following abbreviated training. *Psychonomic Science*, 1968, **10**, 377–378.

SURRIDGE, C. T., & AMSEL, A. Acquisition and extinction under single alternation and random partial-reinforcement conditions with a 24-hour intertrial interval. *Journal of Experimental Psychology*, 1966, **72**, 361–368.

WOODS, P. J., MARKMAN, B. S., LYNCH, W. C., & STOKELY, S. N. Partial reinforcement effects in instrumental escape conditioning. *Learning and Motivation*, 1972, **3**, 279–292.

(Received June 15, 1974)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 484–487

ROTARY ACCELERATION OF A SUBJECT IN[illegible] E REACTION TIME TO MOTION IN PERIPH[illegible]

JAMES M. BORKENHAGEN[1]

California State University, San Jose

The effect of *S*'s rotary acceleration upon choice reaction ti[illegible] accelerating target in peripheral vision was the central issue in [illegible] with the level of acceleration and the viewing angle of the [illegible] variables. Twelve pilots were tested in a rotation device unde[illegible] visual stimulation alone and visual-plus-rotary stimulation. [illegible] sponded to the direction of the visual motion by moving a han[illegible] the right or left. Visual-plus-rotary stimulation produced lon[illegible] than the visual stimulation alone. Choice RT was inversely [illegible] level of acceleration and directly proportional to the viewing [illegible] findings are discussed in connection with theories of double stimula[illegible] intersensory effects.

A number of experiments have demonstrated that reaction time (RT) to a signal can be modified by the presentation of a secondary stimulus in conjunction with the primary signal. Some investigators have found a longer RT to double stimulation, even though the secondary stimulus required no response from the *S* (Koster & Bekker, 1967; Nickerson, 1967). Presentation of stimuli to 2 sensory modalities has, however, produced contrasting results, since intersensory experiments commonly show a facilitation of RT by the accessory stimulus (Bernstein, 1970; Swink, 1966; Symons, 1963). It is clear that either facilitation or inhibition of RT may result from double stimulation, depending upon a number of experimental variables.

The inhibition of RT by accessory stimulation has often been interpreted as evidence for single-channel processing of perceptual information. Theoretical approaches such as those reviewed by Broadbent (1958) emphasize the limited capacity of the central nervous system (CNS) to handle information [illegible]nd, facilitation of RT by a s[illegible]lus has generated anot[illegible]es. Among these are theori[illegible]ming (Symons, 1963) and energy int[illegible]gration (Bernstein, 1970). Explanations of both facilitation and i[illegible] of RT focus upon the informati[illegible] istics of the CNS.

The present study used [illegible] of a visual and a [illegible] stimulus, which were presented to *S* simultaneously. An accelerating visual target served as the visual stimulus, and the vestibular stimulus was a rotary acceleration. A directional response was required to the visual motion only, so the rotary acceleration represented an accessory stimulus. This experiment can therefore be characterized as intersensory, with the 2 stimuli being presented simultaneously and the response required to the visual motion only. The specific purpose of this study was to investigate the effects of 5 levels of rotary acceleration of the *S* on choice RT to the acceleration of a lumin[illegible] line in central and peripheral vision.

Method

Subjects. Twelve commercial airline pilots were paid for their participation in the experiment. All showed normal response to angular acceleration in previous testing.

Apparatus. The Man-Carrying Rotation Device (MCRD) produced the rotary accelerations in the experiment. The MCRD is a 1-degree-of-freedom

[1] This research was conducted at Ames Research Center, Moffett Field, California, under National Aeronautics and Space Administration Grant NGL-05-046-002 to California State University, San Jose. The author thanks Brant Clark, principal investigator, and John Stewart, technical monitor, for their support in this experiment.

[2] Requests for reprints should be sent to James M. Borkenhagen, who is now at the Department of Psychology, University of Southern California, Los Angeles, California 90007.

simulator which rotates the *S* about an earth-vertical axis, providing angular accelerations that have a rise time of about 100 msec. and may be regulated from .01° to 30°/sec². The MCRD has been described in detail by Clark and Stewart (1968). The *S* sat in an aircraft seat at the center of rotation of the MCRD cab. He wore an aviator's helmet, lap belt, and shoulder harness throughout all testing, and he communicated with *E* through a voice communication system. A helmet headrest was provided to hold *S*'s head stationary.

A perimeter arc extended across the horizontal meridian of *S*'s right temporal field of vision. The arc was made of plastic and had a radius of 66 cm. Fixation points consisting of white numerals on a black background were located on the arc at 10° intervals up to 60° from the axis of direct vision. Each numeral (6 × 8 mm.) corresponded to the angular distance from direct viewing, e.g., 0–0°, 1–10°, 2–20°.

The visual stimulus was a movable vertical line (18 cm. long) centered on the screen of a cathode ray tube (CRT). The CRT was located 66 cm. in front of *S*, and its screen measured 36 × 28 cm. The line could be accelerated on the screen to the right or left with a rise time equivalent to that of the rotary acceleration. The *S* responded to the acceleration of the line by deflecting a vertical response stick (8 cm. long) located on the right armrest of the aircraft seat. A force of 110 gm. to the right or 140 gm. to the left was necessary to register a response. The choice RT was registered in milliseconds on a digital voltmeter as the time from the onset of acceleration to the moment *S* responded. An ink-writing recorder collected a record of the accelerations of the line and the MCRD, and the signal from the response stick could also be monitored on this device.

Procedure. The *S* was instructed to respond to the movement of the line by deflecting the response stick in the same direction as the line motion, and he was asked to react as quickly as possible without making an error. When an error was made, the trial was repeated at the end of the series. The *S* sat in the illuminated MCRD cab and viewed the line on the CRT screen. By fixating on the appropriate numeral on the perimeter arc, he observed the line at various points in the left periphery. Binocular vision was used throughout the experiment, but 5 *S*s could not see the line binocularly at 60° because the nasal ridge obscured the vision of the right eye.

There were 2 primary experimental conditions: (*a*) choice reaction to the movement of the line alone, and (*b*) choice reaction to the line motion with simultaneous rotary motion. The rotary acceleration, when present, was in the direction opposite from the acceleration of the line. The 2 accelerations were, however, of equal magnitudes. All *S*s received both stimulation conditions, half of them receiving Condition *a* first and half receiving Condition *b* first. Five levels of acceleration were given in each condition: 1°, 2°, 3°, 6°, and 12°/sec². These levels of angular acceleration are well above the thresholds determined for these pilots by psychophysical methods. Each level of acceleration was tested at the following 7 points in the visual field: 0°, 10°, 20°, 30°, 40°, 50°, and 60° to the left of fixation. The order of presentation of the levels of acceleration and the peripheral angles was randomly determined, as was the direction of acceleration on each trial.

At the start of each trial, the line was centered on the CRT screen. The number of the fixation point for the trial was announced over the voice communication system, and *S* acknowledged the number. Following a standby–ready exchange, the acceleration was applied. The foreperiod from *S*'s reply to the moment of acceleration varied between 1 and 3 sec., so it was not possible for him to anticipate the acceleration with much accuracy. The *S* terminated both the acceleration of the line and the MCRD when he responded, and a minimum of 20 sec. elapsed before the next acceleration.

Three experimental trials were given at each of the 70 combinations of stimulation condition (2), level of acceleration (5), and position in the visual field (7). The testing was organized in blocks, all the trials within a block being at one level of acceleration. Practice trials preceded the experimental trials at all levels of acceleration. Fourteen practice trials were given before the first block of experimental trials in both stimulation conditions: The first 7 trials included all levels of acceleration, whereas the second 7 trials included only the level of acceleration of the experimental trials that followed. The blocks of experimental trials following the first block in each condition were preceded by only 7 practice trials, all of them at the level of acceleration of the subsequent experimental trials. During a testing session, *S* completed 1 or 2 blocks of experimental trials together with the appropriate practice trials. He completed 3 testing sessions in a day: 2 sessions of 2 blocks each and 1 session of 1 block. Two mornings of testing were required of each *S* to complete the experiment.

Results

Choice RT was the dependent variable used in the data analysis. The mean of the 3 experimental choice RTs at each of the 70 treatment combinations was calculated for each *S*, and an analysis of variance was carried out to determine the effects of stimulation condition, level of acceleration, and peripheral angle on choice RT. The analysis followed a 2 × 5 × 7 repeated measures design, with the *S*s factor treated as a random variable. Errors in responding to the direction of the visual acceleration were infrequent, the overall error rate being 1.6%. There were, however, marked differences between *S*s in the number of errors committed.

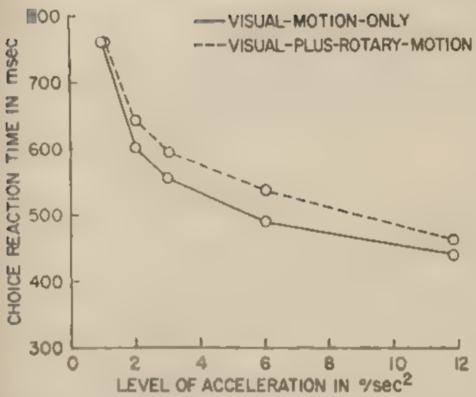


FIGURE 1. Mean choice reaction time at different levels of acceleration for 12 pilots.

Effect of rotary acceleration. The main effect of stimulation condition was significant, F (1, 11) = 11.37, $p < .01$, indicating a significant difference between the choice RT for unisensory stimulation and that for bisensory (see Figure 1). A comparison of the 2 stimulation conditions for each level of acceleration was performed, and the difference between the means of the 2 conditions proved significant in 4 out of 5 levels ($p < .05$). At the lowest level of acceleration (1°/sec²), the difference between the visual-only and visual-plus-rotary motion conditions was not significant ($p > .05$).

Effect of level of acceleration. The significant main effect of level of acceleration, F (4, 44) = 109.35, $p < .001$, demonstrated that the level of acceleration resulted in a decrease in choice RT (Figure 1). It should be noted that in the bisensory condition, the rotary acceleration and visual acceleration levels were both changed together, thus confounding these 2 variables. In spite of this ambiguity, the hypothesis that higher levels of visual acceleration produced lower choice RT is supported by the matching profiles of the bisensory and unisensory data (Figure 1).

Effect of peripheral angle. The main effect of viewing angle was significant, F (6, 66) = 43.13, p < .001, and indicated that there was a significant increase in choice RT as the eccentricity increased (see Figure 2). A subsequent trend analysis confirmed that a significant linear component characterized the trend across the various angles ($p < .01$), and no evidence of a quadratic or higher order component was found ($p > .05$). The significance of the difference between a reference mean (0°) and each treatment mean (10°, 20°, etc.) was tested. All differences were significant ($p < .01$) except for the comparison with the 10° position.

Interaction effects. A significant Levels of Acceleration × Peripheral Angles interaction was found, F (24, 264) = 12.43, $p < .001$, which is visible in Figure 2 as the change in the overall slope of the plots with increased acceleration magnitude. The flattening of the curves with increased level of acceleration indicates that the effect of peripheral viewing is diminished at higher levels of acceleration.

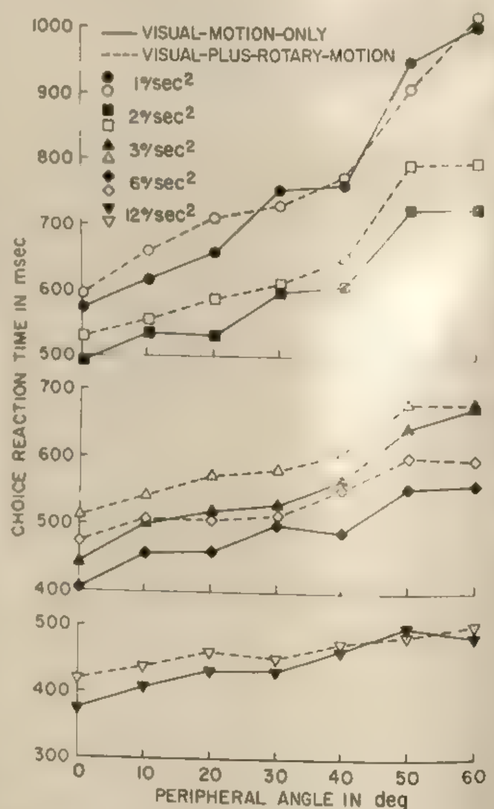


FIGURE 2. Mean choice reaction time at various peripheral angles for 12 pilots.

All other interactions were nonsignificant ($p > .10$).

Discussion

The larger choice RT for bisensory stimulation demonstrates an inhibitory effect of the accessory stimulus upon choice RT. This inhibition of response agrees with the findings of Clark and Stewart[3] for central vision. The absence of a significant inhibitory effect at the 1°/sec² level of acceleration was unexpected, since the low levels of acceleration produced the greatest inhibition in the Clark and Stewart study.

Theories of intersensory facilitation are not applicable to the results of the present experiment; however, the inhibition of response does suggest that some theory of perceptual interference would apply. Response conflict theory (Herman & Kantowitz, 1969) approaches the issue of double stimulation through the relationships of the responses to each stimulus, so an inhibition of response is predicted when there is an opposing response relationship. Because there is a response required to only one of the stimuli in this experiment, the response conflict theory is not properly applicable. Another approach has been concerned with the expectancy of the *S* and the time specification of the stimulus event. This analysis does not appear useful either, since the stimuli were presented simultaneously in this experiment. A popular concept in dealing with the inhibition of response in the double stimulation experiment has been the single-channel mechanism (Welford, 1967). The delay in response to double stimulation is explained by a limitation of the capacity of the CNS to transmit and process information. It may be that the vestibular stimuli are processed to some degree before the visual stimuli are allowed to proceed along the informational pathways. There is also the possibility, however, that a communication channel accepts both stimuli, and handles them simultaneously at a reduced rate of processing.

The choice RT in this experiment was greatly influenced by the viewing angle of the visual stimulus. Sensitivity of the peripheral retina to an accelerating visual target is clearly inferior to the foveal sensitivity, but the difference between peripheral and central sensitivity is diminished by increasing the magnitude of the acceleration. The overall linearity in the pattern of choice RT to an accelerating visual target is consistent with the linearity of sensitivity to a constant velocity stimulus (Leibowitz, Johnson, & Isabelle, 1972).

[3] B. Clark and J. D. Stewart. Effects of angular acceleration on man: Choice reaction time using visual and rotary motion information. Paper presented at the meeting of the Aerospace Medical Association, Las Vegas, Nevada, May 1973. Requests for copies should be sent to Brant Clark, Ames Research Center, Moffett Field, California 94035.

References

Bernstein, I. H. Can we see and hear at the same time? Some recent studies of intersensory facilitation on reaction time. In A. F. Sanders (Ed.), *Attention and performance III.* Amsterdam: North-Holland Publishing Company, 1970.

Broadbent, D. E. *Perception and communication.* London: Pegamon Press, 1958.

Clark, B., & Stewart, J. D. Comparison of sensitivity for the perception of bodily rotation and the oculogyral illusion. *Perception & Psychophysics,* 1968, **3**, 253-256.

Herman, L. M., & Kantowitz, B. H. Effects of second signals occurring after response selection on responses to first signals. *Journal of Experimental Psychology,* 1969, **80**, 570-572.

Koster, W. G., & Bekker, J. A. M. Some experiments on refractoriness. *Acta Psychologica,* 1967, **27**, 64-70.

Leibowitz, H. W., Johnson, C., & Isabelle, E. Peripheral motion detection and refractive error. *Science,* 1972, **177**, 1207-1208.

Nickerson, R. S. Psychological refractory phase and the functional significance of signals. *Journal of Experimental Psychology,* 1967, **73**, 303-312.

Swink, J. R. Intersensory comparisons of reaction time using an electro-pulse tactile stimulus. *Human Factors,* 1966, **8**, 143-145.

Symons, J. R. The effect of various heteromodal stimuli on visual sensitivity. *Quarterly Journal of Experimental Psychology,* 1963, **15**, 243-251.

Welford, A. T. Single channel operation in the brain. *Acta Psychologica,* 1967, **27**, 5-22.

(Received June 25, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 488–493

PROCESSING OF RECENCY ITEMS FOR FREE RECALL[1]

MICHAEL J. WATKINS[2] AND OLGA C. WATKINS

Yale University

The phenomenon of negative recency in secondary memory is usually attributed to the reduced *amount* of rehearsal associated with recency items. On the other hand, negative recency can be explained by the adoption of a different *type* of processing for recency items. An experiment is reported in which the recall of recency items was reduced in an immediate test, but increased in a subsequent test, under conditions in which the recency items could not be identified as such during their presentation. It was concluded that the mode of processing items for an immediate free-recall test is normally modified for recency items.

It is widely held that the serial position curve obtained in single-trial free recall reflects the operation of 2 memory systems. The recency effect is attributed to a fixed-capacity, highly accessible primary memory, while prerecency items are considered to be recalled from a more commodious and permanent secondary memory, with a probability that depends on a wide variety of experimental variables. Upon perception, an item is assumed to enter primary memory, where it remains until displaced. During its stay in primary memory, information concerning the item is transferred to secondary memory. In brief, these are the principal features of the modal model of short-term memory (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965).

This 2-process view of short-term memory was given considerable support by Craik's (1970) finding of a negative recency effect in secondary memory. Craik presented a series of word lists for immediate free recall, and then asked *S*s to recall as many words as possible from all lists. It was found that in this final free-recall test, words presented in recency serial positions were the least well recalled, despite having been recalled well initially. This pattern of results has been replicated in several studies, and has usually been attributed to recency items gaining less rehearsal and therefore less registration in secondary memory than prerecency items. In other words, to the extent that an item is likely to be represented in primary memory at the end of list presentation, the transfer of information about that item to secondary memory will not have run its full course. It follows that items recalled from primary memory in an immediate recall test should be recalled least well on a subsequent recall test, in which only secondary memory is operative.

This quantity-of-rehearsal interpretation of negative recency has been well supported by studies using the overt-rehearsal technique, in which *S* is required to vocalize his rehearsal activity (Rundus & Atkinson, 1970). When the frequency of item rehearsal is plotted against serial position, the last few list items are found to be given progressively less rehearsal, so that the serial position function for rehearsal is very similar to that for the final recall phase of Craik's (1970) procedure (Rundus, 1971). These findings indicate that the number of times an item is repeated correlates strongly with its registration in secondary memory.

However, the notion that secondary memory registration is a simple function of quantity of rehearsal has not gone unquestioned. Even if we accept that the pattern of rehearsal yielded by the overt-repetition technique gives a reasonable indication of rehearsal in more conventional list-presentation procedures, a high correlation between number of rehearsals

[1] This study was partially supported by National Science Foundation Grant GB-24171X to Endel Tulving.

[2] Requests for reprints should be sent to Michael J. Watkins, Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510.

and probability of recall from secondary memory does not prove that repetition is the agent of secondary memory registration. The quantity-of-rehearsal approach has been challenged in a number of recent studies which describe procedures yielding very low or even zero correlations between the amount of rehearsal and secondary memory registration (e.g., Craik & Watkins, 1973; Meunier, Ritz, & Meunier, 1972). Of particular relevance here is an experiment by Craik and Watkins (1973, Experiment 2), in which an initial free-recall test was given immediately after list presentation or after an interval of 20 sec., during which the last 4 items were repeated continuously. A final free-recall test revealed that the last 4 items in the rehearsal condition were no better recalled than those in the control condition. These results were discussed within a depth-of-processing framework (Craik & Lockhart, 1972) with a simple "maintaining" type of rehearsal being assumed during the post-presentation rehearsal interval. A maintaining mode of rehearsal serves to keep items in primary memory, but is ineffective for secondary memory registration.

This approach suggests a quality-of-rehearsal interpretation of the negative recency effect. The negative recency paradigm involves the presentation of a series of lists, with list length typically 12–20 items. It seems reasonable to suppose that *S*s could quickly learn to anticipate the end of the list more or less accurately and so could adopt a maintaining rather than an "elaborative" mode of rehearsal for the last few list items. Thus, the negative recency effect would be explained by recency items being given a qualitatively different type of rehearsal, rather than a reduced amount of rehearsal.

Such a view derives good support from the general lack of negative recency in those studies in which primary memory is eliminated by a distractor task given immediately following list presentation (e.g., Glanzer & Cunitz, 1966). When *S* knows that a distractor task is to follow list presentation, it is to be expected that he will encode the recency items in the same comparatively durable form in which he encodes prerecency items. Jacoby and Bartz (1972) have reported an experiment which clearly favors this interpretation. In brief, they found that the interpolation of a 15-sec. arithmetic task between presentation and initial free recall reduced recency in the initial recall test but, at the same time, eliminated the negative recency effect in final recall. For those subscribing to the quantity-of-rehearsal interpretation of the negative recency effect, however, the fact that the effect is found with Craik's (1970) procedure but not with the distractor paradigm has been something of a paradox. When primary memory is eliminated with a distractor task, there can be no beneficial effects of initial retrieval, so that negative recency should be at least as pronounced as with Craik's procedure.

The aim of the present experiment was to seek further evidence relevant to deciding whether negative recency is better interpreted in terms of type or quantity of rehearsal. The experiment was designed to test a prediction of the quality-of-rehearsal view, a prediction not made by the quantity-of-rehearsal view. The procedure involved presenting word lists of variable length under conditions in which half of the *S*s were given information allowing them to anticipate the end of each list, while list length was unpredictable for the other half of the *S*s. There followed an unexpected final recall test. The quality-of-rehearsal view predicts that, while the final recall of prerecency items will remain unaffected by knowledge of list length, the final recall of recency items will be lower when list length is known. If it is not possible to identify during presentation which are to be recency items, then these items will be afforded the same relatively elaborate processing as prerecency items. On the other hand, when the end of the list can be anticipated, a simple maintenance rehearsal will be adopted for recency items; hence, these items will be recalled with a relatively low probability in the final recall test. A quantity-of-rehearsal view, by contrast, does not predict any difference between the 2 conditions.

Apart from the final recall test, the im-

mediate recall data could also prove of interest. One shortcoming of the quality-of-rehearsal interpretation is that it is post hoc. Recency items are assumed to be given a maintenance type of rehearsal because they are poorly recalled from secondary memory. Why *S*s should change their rehearsal strategy for processing recency items remains an unanswered question. One possibility is that maintenance rehearsal may be better suited to meet the task demands. That is, more recency items may be recalled in the immediate test when a maintenance rather than an elaborative mode of rehearsal is adopted. It should be noted that the evidence currently available lends little support to this possibility. Crowder (1969), using a serial recall procedure, found that knowledge of list length enhanced recall of primacy but not recency items. In addition, a number of studies have shown that the manner in which list items are encoded has little or no effect on the immediate free recall of recency items (e.g., Smith, Barresi, & Gross, 1971). However, the effect of being able to anticipate the end of the list on immediate free recall has so far not been investigated directly.

In sum, the main purpose of the present experiment is to test the hypothesis that recency items will be better registered in secondary memory when they cannot be identified as recency items during their presentation. The experiment also serves to test the possibility that being able to anticipate the end of the list will serve to enhance the recency effect in immediate free recall—a finding which would explain why the mode of rehearsal should be shifted for recency items.

Method

Materials. All *S*s were presented with 7 lists drawn from the same word pool of 98 2-syllable concrete nouns. These lists were of 8, 10, 12, 14, 16, 18, and 20 words. For each group of *S*s, a separate allocation of words to lists and positions was made. This allocation was essentially random, but with restrictions that (*a*) obvious within-lists acoustic and semantic associations were avoided, and (*b*) when measured from the *end* of the list, no word occupied the same list position for any 2 *S* groups.

The *S*s were given booklets for immediate free recall (IFR). Each page was headed with the appropriate list number, and contained a column of ¼-in. squares, drawn down the right-hand edge. For half of the *S*s, the number of cells always equaled the number of words in the list; hence, in this (informed) condition, *S*s knew during presentation which words were to be recency ones. For the other (uninformed) half of the *S*s, the number of cells was always 30.

Subjects and design. A total of 122 undergraduate students took part in the experiment. They were tested in 7 groups, with group size ranging from 12–30 *S*s. An even number of *S*s within each group allowed equal, but otherwise random, *S* assignment to the informed and uninformed conditions. The order of presentation of the 7 list lengths was balanced between groups according to a Latin square arrangement.

Procedure. The *S*s were told that they would be presented with a series of word lists for IFR, and that the number of words per list would vary considerably. The nature of the booklets was explained. As the words of a given list were presented, check marks were to be made in successive cells of the column. The *S*s were informed that if the column lengths varied from page to page, then they were arranged such that list length would always equal the number of column cells. The checking procedure was adopted in order to ensure that *S*s in the informed condition remained fully aware of the stage of list presentation.

The free-recall instructions were slightly modified, in that it was suggested that performance could be maximized by recalling the last few presented items first. This procedure was adopted to ensure that any effect of presentation condition on the IFR of recency items was due to a difference in the availability of primary memory items, and was not merely the result of a difference in retrieval strategy. Presentation was auditory and at a 2-sec. rate. The end of a list was marked by a pencil tap given 2 sec. after the last word. The *S*s in the informed condition were told not to begin recalling between presentation of the last word and the recall signal, but to remain quite still during this interval, in order to avoid giving unintentional cues to the uninformed *S*s. The time allowed for IFR varied between 60 and 90 sec., according to list length. After presentation of the last list a 1-min. arithmetic task was given. The *S*s were then allowed 6 min. for an unexpected final free-recall (FFR) test, in which as many of the presented words as possible were to be recalled.

Results

The main focus of interest will be the effects of knowing list length during presentation on the IFR and FFR of recency items. First, however, the results will be examined at a more general level. To this end, the data were subjected to an analysis

TABLE 1

PERCENT RECALL OF RECENCY ITEMS AS A FUNCTION OF RECALL TEST, PRESENTATION CONDITION, AND LIST LENGTH

Recall test and presentation condition	List length							M
	8	10	12	14	16	18	20	
Immediate								
Informed	65	64	63	66	63	66	67	64.9
Uninformed	60	61	60	61	55	55	57	58.6
difference	5	3	3	5	8	11	10	6.3
Final								
Informed	12	10	12	12	14	13	17	13.4
Uninformed	16	20	16	20	13	15	18	16.8
difference	−4	10	−4	−8	1	2	−1	−3.4

of variance, with recall test (IFR or FFR), presentation condition (whether *S*s were informed or uninformed of list length), position within the list (recency or prerecency) and list length as main effects. The data were collapsed across serial positions within the recency and prerecency components and across individual *S*s within *S* groups, leaving the *S* groups to serve as the random variable. The dependent variable was probability of recall. Recency items were defined as those presented at the last 6 serial positions. The decision to include 6 items in the recency span reflects the fact that this span encompasses the full extent of almost all reports of negative recency, and is also sufficient to capture the greater part of the recency effect in IFR. The statistical inferences to be drawn from the results remain essentially unchanged when 5, 7, or 8 items are used to define the recency span.

The main effects of recall test, position, and list length were all statistically significant, reflecting the higher probability of recall for IFR, recency items, and shorter list lengths. For recall test, $F\ (1, 6) = 1{,}802.26$, $p < .01$; for position, $F\ (1, 6) = 95.62$, $p < .01$; and for list length, $F\ (6, 36) = 7.75$, $p < .01$. Recall was higher in the informed than uninformed condition, but the difference was small and not significant, $F\ (1, 6) = 1.07$, $p > .05$.

Recall test interacted with presentation condition, $F\ (1, 6) = 14.73$, $p < .01$, position, $F\ (1, 6) = 924.72$, $p < .01$, and list length, $F\ (6, 36) = 3.42$, $p < .01$. These effects may be interpreted in terms of the IFR over FFR advantage being greater for the informed than the uninformed condition, for the recency than the prerecency items, and for the shorter list lengths. There was a significant Position × List Length interaction, $F\ (6, 36) = 5.89$, $p < .01$, which was very largely due to the recency over prerecency advantage increasing with list length. Presentation condition did not interact with either position, $F\ (1, 6) < 1$, or list length, $F\ (6, 36) = 1.43$, $p > .05$.

Two second-order interactions proved significant. A Recall Test × Position × List Length effect, $F\ (6, 36) = 2.45$, $p < .05$, was essentially attributable to the decreasing effect of recall test with increasing list length being more pronounced for prerecency than recency items. A Recall Test × Presentation Condition × Position effect, $F\ (1, 6) = 9.83$, $p < .025$, results from opposite effects of presentation condition on the recency items in the 2 tests. Specifically, while knowledge of list length had virtually no effect on prerecency items, it enhanced IFR and depressed FFR of recency items. This interaction is, of course, central to the present study. There is no evidence to suggest that this interaction should be qualified with respect to list length, insofar as the third-order interaction (Recall Test × Presentation Condition × Position × List Length) was not significant, $F\ (6, 36) = 1.56$, $p > .05$.

It will be recalled that the aim of the experiment was to test the hypotheses

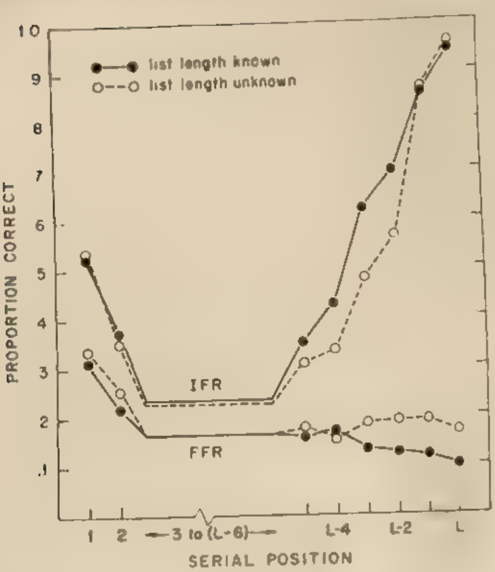


FIGURE 1. Immediate and final free-recall (IFR and FFR) serial position curves for informed and uninformed conditions.

that IFR would be depressed and FFR enhanced for recency items when the end of the list could not be predicted during presentation. While these hypotheses are certainly in good accord with the results of the overall analysis of variance, 2 further analyses were performed in order that they could be tested more directly. The data were collapsed across *S* groups in the first of these tests and across list length in the second.

Table 1 shows recall performance for the informed and uninformed presentation conditions as a function of list length and recall test. In IFR, performance in the uninformed condition was on average only 90.3% of that in the informed condition, whereas in FFR, the difference was in the opposite direction, with recall in the informed condition averaging 80.2% of that in the uninformed condition. Both of these differences were significant: for IFR, t (6) $= 4.94$, $p < .01$, and for FFR, t (6) $= 2.58$, $p < .05$.

Since Table 1 gives no suggestion of a systematic effect of list length on presentation condition, and since the analysis of variance showed that these 2 variables did not significantly interact with each other, either directly or by entering into a significant higher order interaction, it is meaningful to consider the effects of presentation condition with data collapsed across the list length variable. These effects are shown in Figure 1. Each point in the primacy and recency positions is based on 427 observations, while the "asymptotes" of intermediate serial positions are each derived from 2,562 observations. Since the number of observations contributing to the asymptotes varied with list length (from 10 for List Length 8 to 732 for List Length 20), the level of performance for prerecency items should not be taken too seriously. However, Figure 1 does allow a meaningful comparison of the effects of presentation condition on the IFR and FFR within both the recency and prerecency components of the serial position function. It is clear that for prerecency items there is virtually no difference between the presentation conditions. In IFR, the mean probabilities of recalling an item from prerecency positions were .289 and .282 in the informed and uninformed conditions, respectively; in FFR, these probabilities were .190 and .197. In contrast to the prerecency items, the recency items were appreciably affected by presentation condition. With scores collapsed over serial position, the superior IFR for *S*s in the informed condition was significant, t (120) $= 4.27$, $p < .01$, as was the superior FFR of the uninformed condition, t (120) $= 3.01$, $p < .01$.

DISCUSSION

The principal aim of the experiment was to test the prediction that in a delayed (FFR) test, recall of recency items would be enhanced when the end of the list could not be anticipated during presentation. This prediction was confirmed, and hence the notion that the negative recency effect results from switching to a different mode of processing for recency items was supported. It is not immediately apparent how the effect of presentation condition on FFR can be reconciled with the view that negative recency is due to the last few list items being given a progressively decreasing amount of a single rehearsal process (Rundus, 1971).

The quantity-of-rehearsal view has similar difficulties in accounting for the effect of presentation condition on the IFR of recency items. On the other hand, the superior IFR recency effect shown in the informed condition is clearly in line with the notion that mode of processing is normally switched for recency items. In addition, the IFR results have more general theoretical implications. There is an extensive literature documenting the effects of a wide range of variables to be confined to prerecency serial positions (for a summary, see Glanzer, 1972). To date, only presentation modality has been unequivocally shown to have the converse effect, with an auditory over visual advantage restricted to recency items (Murdock & Walker, 1969). Moreover, it is by no means agreed that this modality effect qualifies the primary memory property of invariant capacity, in that the auditory advantage is often attributed to a separate memory system (an interpretation that is perhaps less readily applied to the present findings). In this connection it should be noted that the effects of modality and of being able to anticipate the end of the list are not identical. Unlike the modality effect, the effect of knowing when to expect the end of the list is not apparent at the last 1 or 2 serial positions, but is observed only over the earlier portion of the recency effect. Although it is not entirely clear why the effect of presentation condition should be so specific with respect to serial position, it may be speculated that when *S* believes that the end of the list is imminent, he switches to a mode processing somewhat similar to that used in a memory span task.

We have noted that the quantity-of-rehearsal account of secondary memory registration does not predict the effects of presentation condition in either recall test. A far greater difficulty for this account lies in the fact that the effects were in opposite directions in the 2 recall tests. On the other hand, the interaction between recall test and presentation condition is not difficult to understand when qualitatively different types of rehearsal are allowed. Specifically, when a list of verbal items is presented for immediate free recall, a certain amount of processing capacity will be reserved for the elaborative and organizing operations which are collectively referred to as secondary memory registration. When the end of the list is thought imminent, secondary memory registration is reduced in order to increase primary memory capacity. The result is that recall of recency items is maximized in an immediate test, but at the expense of being reduced in a subsequent test in which only secondary memory is operative.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 2. New York: Academic Press, 1968.

Craik, F. I. M. The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 143–148.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671–684.

Craik, F. I. M., & Watkins, M. J. The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, in press.

Crowder, R. G. Behavioral strategies in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 524–528.

Glanzer, M. Storage mechanisms in recall. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 5. New York: Academic Press, 1972.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

Jacoby, L. L., & Bartz, W. H. Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 561–565.

Meunier, G. F., Ritz, D., & Meunier, J. A. Rehearsal of individual items in short-term memory. *Journal of Experimental Psychology*, 1972, **95**, 465–467.

Murdock, B. B., Jr., & Walker, K. D. Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 665–676.

Rundus, D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 1971, **89**, 63–77.

Rundus, D., & Atkinson, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 684–688.

Smith, E. E., Barresi, J., & Gross, A. E. Imaginal versus verbal coding and the primary-secondary memory distinction. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 597–603.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

(Received June 26, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 494–497

REACTIONS TO FRUSTRATIVE NONREWARD AS A FUNCTION OF PERCEIVED LOCUS OF CONTROL OF REINFORCEMENT

J. WESLEY LIBB[1] AND CAMELLA SERUM

University of Alabama

Based on Rotter's definition of the personality variable locus of control of reinforcement—it was predicted that internal *Ss* in an ambiguous experimental situation would be responsive to frustrative nonreward whereas external *Ss* might not be similarly frustrated. As predicted, internal *Ss* responded with reduced latencies following nonreward and, early in learning, made significantly more errors on a subsequent complex task. Consistent with Spence's theory of emotionally based drive, these same *Ss* showed improved performance following frustration at a later stage in learning when the task was largely mastered. No similar experimental effects emerged for the external *Ss* suggesting that only the internal *Ss* were significantly frustrated by the experimental procedures.

The hypothesis that people with differing personality characteristics react differently to frustrating situations is not a unique or novel hypothesis. Neither is the postulation that differing experiences may be differentially frustrating to various individuals. The possibility, however, of being able to predict in what situations a certain individual might be frustrated and the effects of his being frustrated upon his behavior in various situations would seem a meaningful contribution to attempts to relate learning theory to human behavior.

According to Amsel's (1958) theory of frustrative nonreward, failure of reward following the establishment of an expectancy for reward elicits a primary frustration reaction, R_f. Feedback from this reaction leads to enhanced motivational effects evident in subsequent instrumental behavior. The alleged motivational effects are typically evident in the invigoration of responses following nonreward. The term *frustration effect* (FE) refers to this invigoration of performance. Numerous studies attest to the empirical validity of frustration theory in regard to FE with both animals and children (Ryan & Watson, 1968; Wagner, 1959). The data for adult human *Ss* are lacking, with one recent exception (Libb, 1972), perhaps due to difficulties encountered in providing material incentives outweighing *Ss*' self-induced symbolic motivations.

The FE has generally been demonstrated in a task where there are few obvious competing responses and with a response well established by previous training. In general, frustration has resulted in increased running speeds and decreased latencies, which may be said to "enhance" performance. On more complex tasks, in which there are more obvious competing responses, effects differ. Both Schmeck and Bruning (1968) and Libb (1972) lend support to the hypothesis that frustration produces impaired performance on a complex task. The relationship between frustration and performance is not a simple, positive linear function. Based on Spence's (1960) theory of emotionally based drive, Libb predicted that during the early stages of learning a complex task, more errors would occur following a frustrating condition than following a neutral or rewarding situation. In later stages of learning, it was predicted that frustration would be less detrimental and might be associated with fewer errors, i.e., an interaction over trial blocks was predicted. These hypotheses were largely supported.

While FE is a reliable experimental phenomenon, it has been noted by several authors that individual differences may occur in reaction to frustrative nonreward in human research. Ryan and Moffitt

[1] Requests for reprints should be sent to J. Wesley Libb, Ridgecrest Children's Center, University of Alabama, Box 2232, University, Alabama 35486.

(1966) found that only approximately half of their *S*s demonstrated FE. They argued that this differential reaction may not be an isolated chance occurrence but rather an individual difference variable generally obscured by the use of averages. Little experimental attention has been devoted to systematic analysis of individual differences in regard to frustration theory.

Ryan and Watson (1968) suggested that the individual difference variable, locus of control of reinforcement (Rotter, 1966), may account for differential responding to nonreward. Internal-external locus of control of reinforcement refers to the extent to which a person perceives reinforcement as under his own control (internal), or under the control of powerful others, fate, or chance (external).

In the research that follows, a design similar to that of Libb (1972) was employed. Human *S*s were given ambiguous instructions with regard to locus of control of reinforcement. Based on previous research findings (Rotter, Liverant, & Crowne, 1961), it was assumed that internal *S*s would develop a higher expectancy for reward than external *S*s and would be more responsive to nonreward than externally oriented *S*s. Specifically, it was predicted that internal *S*s would show a significant reduction in response latency following nonrewarded trials. Should external *S*s manifest this effect at all, it was hypothesized to be of lesser degree than for internal *S*s. With regard to errors for the internal group, it was predicted that in the early stages of learning frustration would be evident in a decrement in performance or as increased errors following nonreward. Later in learning this effect was anticipated to disappear or a decrease in errors following nonrewarded trials actually would be evident. Since the external group was predicted to be less frustrated than the internal group, these effects were predicted to be of lesser magnitude, if present at all, for the external group.

Method

Subjects. Thirty-two *S*s from introductory psychology classes at the University of Alabama participated as part of a course requirement. All 32 *S*s volunteered to participate with no prior understanding of the precise nature of the experiment. Rotter's (1966) Internal-External Control Scale was administered to all *S*s who were divided into 2 groups based on a median score of 10.5. The internal group consisted of 16 internal *S*s (8 male and 8 female) and the external group, 16 external *S*s (11 male and 5 female). Little effort was made to match sexes within groups inasmuch as previous literature reveals no sex differences. Also, a *t* test for sex differences within this experiment was not significant.

Apparatus. The experimental chamber was a sound-dampened 10 × 10 ft. (3 × 3 m.) cubicle. Electromechanical programming equipment was located in a separate room. Manipulanda and the reward dispenser were mounted in the experimental chamber. Directly above a small table a Gerbrands Universal Feeder was mounted from which marbles were dispensed for appropriate responses in a 2-task sequence. The first manipulandum in position on the table was a single button (1 in. [2.54 cm.] in diameter) mounted on an 8 × 10 in. (2 × 2.5 dm. in diameter) panel. Immediately above the button on the panel a small 28-w. stimulus light was mounted. The second panel, 8 × 14 in. (2 × 3.5 dm.) in size, separately mounted to the right of the first, consisted of 4 buttons centrally positioned below a multiple-stimulus projector (Grason-Stadler, No. 4580-159), also mounted on the panel. Complexity on the second task, as part of a 2-task sequence paralleling Amsel's (1958) apparatus, was introduced by varying the stimuli presented on the multiple-stimulus projector which cued *S* as to the correct button on that trial.

Procedure. The *S*s participated to earn a predetermined number of marbles to qualify for extra credit in course work. They were escorted to the experimental chamber and seated before the apparatus. The *E* read the following directions:

> You have fifteen minutes to earn 46 marbles to qualify for course credit in this experiment. Marbles will be earned by pressing the button when the light comes on. Press the button on the first panel when the light comes on and the first button on the second panel when the colored light comes on. After you earn 10 marbles, there will be a slight delay and change in the program.

Instructions were repeated if procedure clarification was requested by *S*. All *S*s performed 5 training trials and 24 experimental trials. A trial consisted of a sequence of responses on both panels. The sequence started with illumination of the stimulus on the first panel. Five seconds after *S* responded on this panel, the multiple-stimulus projector on the second panel was illuminated permitting *S* to choose 1 of 4 buttons. The sequence recycled 30 sec. after the appropriate response on the second panel. Reinforcement consisted of the delivery of one marble on rewarded trials. On training trials for the second task only one stimulus was presented and the first button was always the correct response.

TABLE 1

Mean Errors and Latencies per Trial Block

Group and condition	Errors			Latency		
	Block 1	Block 2	Total	Block 1	Block 2	Total
Internal						
Nonrewarded	9.44	3.69	6.57	1.41	1.26	1.34
Rewarded	6.56	5.13	5.85	1.82	1.48	1.65
External						
Nonrewarded	7.50	3.69	5.60	1.74	1.72	1.73
Rewarded	8.00	2.75	5.38	1.81	1.68	1.75

Following training trials, stimuli on the projector were scrambled and *S* was required to learn the appropriate button to be paired with red, blue, yellow, or green lights. On this panel, *S* could press as many buttons as he wished as many times as he wished, but reward was forthcoming only when the correct button was pressed. Marble reinforcement for correct responses was continuous on the second panel for all *S*s. Both groups received randomized 50% reinforcement on Task 1 during experimental trials, with the one limitation that no more than 2 trials of either reward or nonreward occurred in succession. In addition to the color presented on the multiple-stimulus projector, irrelevant cues (an X, a square, a circle, or a dash) were also presented simultaneously so as to enhance the complexity of the task.

For each rewarded and nonrewarded response on Task 1, number of incorrect responses and latency to initial response on Task 2 were recorded. Latency was defined as the time from illumination of the multiple-stimulus projector to *S*'s first button press on Panel 2, whether correct or incorrect.

Results

The latency scores were calculated for each *S* following both rewarded and nonrewarded trials in Blocks 1 and 2 (see Table 1). Since the homogeneity of variance assumptions could not be established for these 2 groups, $F_{max} = 1.977$, $p < .01$, latency scores for each of the 2 groups were analyzed separately with a Treatment × Treatment × *S* analysis of variance. The internal group demonstrated a significant main effect, $F\ (1,\ 15) = 4.282$, $p < .05$, while the external group did not exhibit this effect. The internal group thus demonstrated significantly reduced latencies following nonrewarded trials when compared with rewarded trials.

Four separate error scores were calculated for each *S*: total error scores following nonreward and following reward in each of 2 blocks of experimental trials (see Table 1). A 3-factor analysis of variance on these data yielded a significant triple interaction, $F\ (1,\ 30) = 6.38$, $p < .02$, and a significant trial blocks effect, $F\ (1,\ 30) = 50.962$, $p < .001$. The latter indicates that *S*s performed with significantly fewer errors in the second block of trials than the first block. A Treatment × Treatment × *S* analysis of variance was performed separately on each of the 2 groups. A significant A × B interaction with the internal group indicated the appropriateness of individual *t* tests, $F\ (1,\ 15) = 5.684$, $p < .025$. Internal group *S*s made significantly more errors following nonreward than following reward in Block 1, $t\ (15) = 4.496$, $p < .01$. Although the *F* test was statistically significant for the interaction, the *t* test for the second block of trials on error scores was not statistically significant, $t\ (15) = 1.118$, $p < .3$. Again the trial blocks effect was statistically significant, $F\ (1,\ 15) = 15.788$, $p < .005$. The Treatment × Treatment × *S* analysis of variance for external group *S*s yielded only a trial blocks effect, $F\ (1,\ 15) = 43.01$, $p < .001$.

Discussion

These data, both error and latency scores, suggest that the internal group was frustrated by the experimental manipulation whereas the external group was not similarly frustrated. Hypotheses were thus supported in that the internal group showed a significant latency effect consistent with the animal literature and a significant interaction with regard to error scores, consistent with Libb (1972). The interaction was clearer than in Libb's study, perhaps due to the fact that *S*s who had little tendency to be frustrated by experimental manipulations were classified as externals. This experiment suggests that much of the difficulty in demonstrating FEs with human *S*s may be due to the fact that individual differences obscure results when such factors are undetected and uncontrolled. If, in fact, an *S* is not frustrated, it hardly seems reasonable to predict that he would manifest FE.

It is also noteworthy that the latency effect was observed in this experiment and not observed in the Libb (1972) experiment in which

both internal and external *S*s were likely participants. These data are thus unique in that the FE was evident in 2 dependent variables: latency and error scores. Such findings lend further support to Spence's (1960) theory of emotionally based drive and its effect on performance of tasks of varying complexity.

Difficulties in interpretation may be encountered in the lack of a significant reduction in errors from Block 1 to Block 2 following reward in the internal group. Since trials introducing both nonreward and reward were randomly but evenly distributed within each block, one would expect errors to decrease following both reward and nonreward as learning progressed. The internal *S* may have become bored with the task such that his performance deteriorated unless frustration was added to the motivational complex.

Alternative explanations of these data are possible. For example, internal *S*s more than external *S*s may initially be searching for a combination Task 1–Task 2 strategy which will permit them to be correct 100% of the time. Such an approach to the experimental situation could account for not only the higher number of errors following nonreward but also the equivalent reward performance on Blocks 1 and 2; i.e., occasionally trying out more complex strategies intrudes on performance following rewarded trials. Alternatively internal *S*s may have experienced some anxiety with regard to Task 1 since the nature of the task may have precluded their preferred perception of skill control. Increased anxiety may lead to enhanced drive and increased errors in parallel to the predicted FEs.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

LIBB, J. W. Frustration and task complexity: An extension of frustration theory. *Journal of Experimental Psychology*, 1972, **96**, 67–72.

ROTTER, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 1966, **80**(1, Whole No. 609).

ROTTER, J. B., LIVERANT, S., & CROWNE, D. P. The growth and extinction of expectancies in chance controlled and skilled tasks. *Journal of Psychology*, 1961, **52**, 161–177.

RYAN, T. J., & MOFFITT, A. R. Response speed as a function of age, incentive value, and reinforcement schedule. *Child Development*, 1966, **37**, 103–113.

RYAN, T. J., & WATSON, P. Frustrative nonreward theory applied to children's behavior. *Psychological Bulletin*, 1968, **69**, 111–125.

SCHMECK, R. R., & BRUNING, J. L. Task difficulty and the frustration effect. *Journal of Experimental Psychology*, 1968, **78**, 516–520.

SPENCE, K. W. *Behavior theory and learning: Selected papers*. Englewood Cliffs, N.J.: Prentice-Hall, 1960.

WAGNER, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *Journal of Experimental Psychology*, 1959, **57**, 130–136.

(Received May 26, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 498–502

METACONTRAST AND SIMPLE REACTION TIME: A REEXAMINATION [1]

ROBERT W. PROCTOR AND IRA H. BERNSTEIN [2]

University of Texas at Arlington

D. L. SCHURMAN

Emory University

Fehrer and Raab and subsequent investigators have reported that simple reaction time (RT) to a metacontrast-inducing combination of test and mask stimuli is more rapid than to either component (test or mask) alone. Thus, they concluded that simple RT is not affected by metacontrast. We note that their conclusion is not warranted. A test–mask combination normally should evoke faster RT than either test or mask alone because of the greater area and, hence, total energy of the combination. A more appropriate comparison is between the metacontrast-inducing combination and a combination of equal area and energy that does not induce metacontrast. The results of this comparison were that 4 *S*s gave equivocal results; RT to the 2 combinations and to its components were all equal, but 4 *S*s showed faster RT to the non-metacontrast-inducing combination. The overall results indicated that metacontrast does, in fact, affect simple RT.

Metacontrast refers to a decrement in the clarity or visibility of a stimulus produced by adjacent, nonoverlapping stimuli that occur in close temporal proximity. One typical configuration used in studies of metacontrast is a centrally located square (test stimulus) and a pair of flanking squares (mask stimuli) which share adjacent contours with the test stimulus. The test, mask, and combination of test and mask occur in random order, over trials. The temporal spacing between test and mask—stimulus onset asynchronies (SOAs)—are generally varied over blocks of trials. When the test and mask energies are equal, the phenomenal brightness of the test square is a U-shaped function of SOA. With test energy less than mask energy, phenomenal brightness is an increasing function of SOA. In either case, the phenomenal brightness of a test stimulus is reduced by the presence of nonoverlapping mask stimulus. (See Kahneman, 1968, for a general substantiative and procedural review.)

Despite the phenomenal dimming of the test stimulus, however, simple reaction time (RT) to a test–mask combination is typically no slower than simple RT to the test alone. This finding has led several investigators to conclude that metacontrast does not affect simple RT (Fehrer & Raab, 1962: Harrison & Fox, 1966; Schiller & Smith, 1966). Fehrer and Raab further concluded that "It was therefore the physical dimensions of the [test] stimulus rather than its phenomenal characteristics that determined RT [p. 147]."

However, the support for the hypothesis that simple RT and verbal report obey different functional relationships may simply be based upon an improper comparison. The RT to the combination should not be compared to the test alone. A more proper comparison is with a second combination of equal energy and area which does not produce metacontrast. Such a combination could consist of squares of the same individual dimensions, durations, and luminances as the metacontrast combination but which are spatially arrayed so as not to induce metacontrast (i.e., are nonadjacent).

[1] This research was supported by the Liberal Arts Organized Research Fund, University of Texas at Arlington.

[2] Requests for reprints should be sent to Ira H. Bernstein, Department of Psychology, University of Texas, Arlington, Texas 76010.

Comparing simple RT to a metacontrast-inducing (*experimental*) combination with a non-metacontrast-inducing (*control*) combination rather than the mask alone is necessary to equate for stimulus energy. The RT to both combinations can reflect spatial integration and, when SOA is not zero, temporal integration as well (Bernstein, Amundson, & Schurman, 1973; Kietzman & Gillam, 1972). Spatial integration was reported by Fehrer and Raab (1962) among many others; the 2 mask squares, being of greater combined area, produced faster RT than the single test square. The crucial question is whether metacontrast does or does not affect integrative processes.

The preceding implies that 2 additional findings must converge upon the observation that RT to an experimental combination is faster than RT to the mask alone (and, consequently, the test alone as the greater mask area normally yields the shorter RT of the 2 conditions) before one can conclude that metacontrast does not affect simple RT. One should demonstrate that the experimental configuration did, in fact, produce metacontrast as inferred from verbal report. In addition, RT to the experimental and control combinations should be equal if the extent of integration is the same. Conversely, if RT to the control combination is faster than RT to the experimental combination, one should conclude that metacontrast does affect RT even if RT to both combinations is faster than RT to the mask alone.

A third possible result is that RT to the mask and the 2 combinations are all equal to one another. This finding would be uninformative, since it would indicate that the mask alone was the effective stimulus, i.e., the additional energy provided by the test was not sufficiently great to affect RT.

The present data were gathered from the third study of a series. The first 2 studies employed a less optimal design and will be discussed briefly. As with other metacontrast studies in which the mask energy exceeded the test energy, pilot data revealed that the verbally inferred masking function for the experimental combination was monotonic. Since the metacontrast effect was greatest with concurrent test–mask onset, 0-msec. SOA was best suited to explore possible effects of metacontrast upon simple RT. Hence, SOA was not varied in the experiment reported here although some relevant data were gathered in the earlier experiments.

Method

Subjects. The *S*s were 8 paid volunteers, 4 males (RF, JSA, GB, and KS) and 4 females (VA, JS, SF, JC). All were graduate students in psychology at the University of Texas at Arlington. All had normal or corrected-to-normal visual acuity.

Apparatus and stimuli. Stimuli were presented by transillumination in a 3-channel tachistoscope (Scientific Prototype, Model GB) using standard Sylvania bulbs (F8T5/D). Because imperfections in mirror alignment might have produced binocular parallax, viewing was monocular through a 2-mm. artificial pupil. Luminances were measured with a McBeth Illuminometer. The tachistoscope's timing accuracy was verified with an oscilloscope (Tektronix, Model 564). The RT responses were right-hand telegraph key depressions timed with a Hunter Klockounter (Model 120G). A foot-pedal was used for *S* to self-initiate trials with the left foot. The *E* and the *S* were seated in adjacent rooms.

The stimulus display consisted of a dim fixation circle that remained on continuously. The circle was located in the blank field of the tachistoscope and was illuminated by an external source so that the blank field could be used as a separate exposure field. The fixation point had very low luminance and became clearly visible only after fixation. Exposure Field 1 contained a centrally located square, the base of which was .33° of arc above fixation (experimental test.) Exposure Field 2 also contained a centrally located square, the top of which was .33° of arc below fixation (control test). Exposure Field 3 (nominally the blank field) contained a pair of squares (the mask), whose inner sides were contiguous to the experimental test. The individual squares were all transilluminated. The dimensions (.67° of arc/side) and luminances (63 cd/m²) were identical to those used by Fehrer and Raab (1962).

The test and mask durations were 2 and 100 msec., respectively. This typically produced chance accuracy in discriminating experimental test + mask (TM–E) from mask (M) alone at an SOA of 0, the SOA used throughout the experiment. This duration was adequate to allow perfect discrimination of control test + mask (TM–C) from M alone in pilot work. The entire display fell within a 2° arc of fixation.

Procedure. All but 2 *S*s had extensive practice at reacting to the various stimulus conditions in either or both of 2 pilot studies and were run without additional practice. The 2 experimentally naive *S*s were run for a 1-hr practice period. All *S*s were then run in a single session of approximately 1 hr.

The session began with 5 min. of dark adaptation. Half of the RT trials were then run. A 3-min. break was provided following which a series of verbal report trials, described below, were run. A second 3-min. break was provided, and the remaining RT trials were run.

Both verbal report and RT trials began with a signal from *E* that the necessary adjustments had been made. The *S* then fixated, blinked if necessary, and when the circle appeared as sharp and clear as possible, initiated the trial with the foot switch. This activated a timer to initiate 1 of 3 randomly chosen foreperiods (1.9, 2.2, or 2.5 sec.).

For RT trials, 1 of 5 stimulus conditions was possible: (*a*) TM–E, (*b*) TM–C, (*c*) M alone, (*d*) experimental test (T–E) alone, and (*e*) control test (T–C) alone. These occurred randomly, subject to the condition that TM–E, TM–C, and M occur 20 times each and T–E and T–C occur 10 times each per half session. A complete factorial design combining the 2 test positions with masks above and below fixation was not possible since there was no fourth tachistoscopic field available for a mask below fixation. Thus, one purpose of the T–E vs. T–C comparison was to insure that the retinal regions stimulated were equally sensitive.

All responses were timed from the end of the foreperiod. All RTs longer than 350 msec. or shorter than 100 msec. were rerun (less than 1% of all trials).

The verbal report trials consisted of the random presentation of M and TM–E 50 times each. The *S*s responded on each of these trials along a 6-point confidence rating scale (Green & Swets, 1966), with 1 = very sure that TM–E had been presented, 2 = moderately sure that TM–E had been presented, . . . and 6 = very sure that M had been presented. As noted above, pilot data indicated that the TM–C and M could always be discriminated from one another. Likewise, T–C and T–E were never confused with each other or with the other alternatives.

Results and Discussion

The major finding was that metacontrast did affect simple RT. The RT to TM–C (201 msec. over *S*s) was faster than RT to TM–E and RT to M which, in turn, did not differ (205 msec.). The RT to T–E and T–C, likewise, did not differ, and due to their smaller area were the slowest (216 msec.). Analysis of variance (Stimulus Conditions × *S*s) indicated that the 5 means differed, F (4, 28) = 24.13, $p < .01$. Also, RT to TM–C was significantly faster than RT to TM–E, F (1, 7) = 5.67, $p < .05$, but RT to M and TM–E did not differ ($F < 1.0$) likewise, RT to T–E and RT to T–C did not differ ($F < 1.0$).

Receiver operating characteristic (ROC) curves were obtained from each *S*'s confidence ratings, treating TM–E presentations as "signal plus noise" and M presentations as "noise." The mean area under these ROC curves was .524. This is well within chance limits, indicating that concurrent presentation of the experimental test (T–E) and the mask (M) produced complete masking. Only 1 *S* (KS) exceeded chance.

Thus, the data show that more integration of test and mask energy occurred without rather than with metacontrast. In fact, with the present high level of metacontrast, no integration at all occurred; RT to the experimental combination and mask alone were equal. Fehrer and Raab's (1962) conclusion that metacontrast does not affect RT seems to be an artifact of

TABLE 1

Individual S Mean ($\bar{X}$) Reaction Time (RT) and Standard Deviations (*SD*s) as a Function of Stimulus Condition

Subject	Stimulus condition: TM–E	TM–C	M	T–E	T–C
GB					
$\bar{X}$	221	222	222	237	240
SD	18	20	16	23	20
JS					
$\bar{X}$	177	177	177	196	185
SD	13	11	16	20	13
RJ					
$\bar{X}$	211	211	208	218	217
SD	20	18	18	21	15
SF					
$\bar{X}$	220	219	218	225	236
SD	24	21	27	19	22
KS					
$\bar{X}$	187	180[a,*]	189	209	203
SD	22	25	25	27	33
JSA					
$\bar{X}$	224	211[b,**]	225	228	231
SD	19	24	23	32	29
JC					
$\bar{X}$	209	201[b,*]	203	212	215
SD	21	18	25	24	25
VA					
$\bar{X}$	191	186[a,**]	195	203	200
SD	20	19	21	14	27
Comp.					
$\bar{X}$	205	201	205	216	216

Note. Abbreviations: TM–E = experimental test + mask, TM–C = control test + mask, M = mask alone, T–E = experimental test alone, and T–C = control test alone. All tests are 1-tailed.

[a] RT to TM–C < RT to M.

[b] RT to TM–C < RT to TM–E.

* $p < .05$.

** $p < .01$.

improperly comparing RT to the experimental combination with RT to the mask instead of RT to the control combination. Our results are not due to differential sensitivity of the retinal loci stimulated by the 2 test squares, since RT to each of them was identical.

The individual *S* data are presented in Table 1. The results of *t* tests comparing TM–C, TM–E, and M among themselves are included. Only 4 *S*s (KS, JSA, JC, and VA) had faster RTs to the control combination than to the experimental combination and mask alone. The remaining *S*s' RTs to these 3 stimulus combinations are all equal; hence, their data are inconclusive. The test energy was apparently too low for the latter *S*s to demonstrate any integration. No *S* fulfilled the criteria discussed earlier to support Fehrer and Raab's (1962) conclusion by manifesting equal amounts of integration to the 2 combinations. Since enough *S*s were run in the present study to demonstrate the metacontrast impairment in RT statistically over all *S*s, there seemed to be no reason to run additional trials or *S*s. We did look for factors separating the 2 groups of *S*s who, respectively, did and did not demonstrate integration with T–C. These analyses failed to show that verbal report performance, familiarity with the metacontrast phenomenon, RT experience, or *S*'s own expectancies were pertinent.

The results of our 2 prior studies (Bernstein, Amundson, & Schurman, 1973; Bernstein, Futch, & Schurman, 1973) were that RT to TM–E fell between the RT to TM–C and RT to M, unlike the present results. This was because the test energy was sufficiently high to allow partial verbal report detection and, hence, integration. The stimulus parameters in the Fehrer and Raab (1962) and related studies suggest partial detection. This would explain why RT to TM–E would be faster than RT to M. Despite the energy differences among the studies, the difference at 0-msec. SOA between RT to TM–C and RT to TM–E was about the same 4 msec. as shown here. These earlier studies also found that not all *S*s demonstrated integration effects.

The failure to obtain integration effects with the present high levels of metacontrast is contrary to findings reported by Fehrer and Biederman (1962) when (unlike the Fehrer & Raab 1962 study) verbal report performance was near chance due to a slightly different method of presenting the test. Fehrer and Biederman reported integration did occur with chance verbal report accuracy, a form of "discrimination without awareness." Recently, Bernstein, Admunson, and Schurman (1973) reported an artifact underlying Fehrer and Biederman's conclusion. The artifact was a difference of statistical power between the verbal report comparison and the RT comparison. Bernstein, Amundson, and Schurman 1973 used signal detection methods to compare verbal report and RT directly, using inferential tests of similar power. They found verbal report to be the superior performance index.

The results presented in this report and those of Bernstein, Amundson, and Schurman (1973) jointly suggest that verbal report and RT do not differ as much in the *type* of perceptual mechanism but, rather, in the *duration* of processing that occurs prior to a response. That is, a verbal discrimination, or more generally, any response requiring a choice, involves a longer period of temporal integration within the system. Additional evidence for this difference in processing duration is the shorter critical duration for simple RT when compared to verbal report (Bernstein, Futch, & Schurman, 1973; Bruder & Kietzman, 1973; Fehrer & Raab, 1962; Kietzman & Gillam, 1972). It is with this observation, first made in the Fehrer–Raab series of experiments and their most important conclusion, that we strongly agree. The shorter period of temporal integration would explain why Fehrer and Raab and others, including us in our pilot studies, found no difference between RT to the combination and RT to the mask alone at longer SOA even though verbal report was impaired.

Special note needs to be made about Fehrer and Raab's (1962) 25–75 msec. range of SOA, where there was partial to total phenomenal test suppression. The RT to the combination was essentially equal to the RT to the test (but slower than RT to the mask). As the authors noted,

*S*s easily learned to discriminate the combination from the mask by use of apparent movement cues. They also noted that these cues could have constituted part of the effective stimulus for RT to the combination. Thus, the near equality of RT to the combination and test alone seems to reflect opposing processes; apparent movement decreased RT to the combination whereas phenomenal dimming increased it. The surprising point of their experiment thus does not seem to be that these RTs were essentially equal. Rather, it is why the second event (mask) could influence simple RT at SOAs beyond the critical duration for the first event (test). There are abundant data to suggest that lengthening the duration of a target of moderate to high luninance beyond 10–20 msec. does not reduce and may even increase simple RT (Bernstein, Futch, & Schurman, 1973). Bernstein, Futch, & Schurman argued for the importance of "on" and "off" responses as RT determinants. These effects also seem appropriate to consider in the light of the present paradox.

Thus, one need not assume that the physical stimulus intensity is processed at one level, and that metacontrast affects a higher, more conscious level. Mechanisms, such as lateral inhibition and luminance summation, which collectively seem to account for much of the metacontrast phenomenon, can occur at the earliest stages of information processing and thus affect both verbal report and RT.

REFERENCES

Bernstein, I. H., Amundson, V. E., & Schurman, D. L. Metacontrast inferred from reaction time and verbal report: Replication and comments on the Fehrer–Biederman experiment. *Journal of Experimental Psychology*, 1973, **100**, 195–201.

Bernstein, I. H., Futch, D. G., & Schurman, D. L. Some exposure duration effects in simple reaction time. *Journal of Experimental Psychology*, 1973, **97**, 317–322.

Bruder, G., & Kietzman, M. L. Visual temporal integration for threshold, signal detectability and reaction time measures. *Perception & Psychophysics*, 1973, **13**, 293–300.

Fehrer, E., & Biederman, I. A comparison of reaction time and verbal report in the detection of masked stimuli. *Journal of Experimental Psychology*, 1962, **64**, 126–130.

Fehrer, E., & Raab, D. Reaction time to stimuli masked by metacontrast. *Journal of Experimental Psychology*, 1962, **63**, 143–147.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Harrison, K., & Fox, R. Replication of reaction time to stimuli masked by metacontrast. *Journal of Experimental Psychology*, 1966, **71**, 162–163.

Kahneman, D. Method, findings, and theory in the studies of visual masking. *Psychological Bulletin*, 1968, **70**, 404–425.

Kietzman, M. L., & Gillam, B. J. Visual temporal integration and simple reaction time. *Perception & Psychophysics*, 1972, **11**, 333–340.

Schiller, P. H., & Smith, M. C. Detection in metacontrast. *Journal of Experimental Psychology*, 1966, **71**, 32–39.

(Received July 9, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 503-507

RELATIVE EFFECTIVENESS OF RHYMES AND SYNONYMS AS RETRIEVAL CUES[1]

DOUGLAS L. NELSON[2] AND DAVID H. BROOKS

University of South Florida

An experiment was performed to compare the relative effectiveness of rhymes and synonyms as extralist retrieval cues. The a priori relation between synonym cues and their targets and rhyme cues and their targets was equated, at both weak and strong levels, by the use of normative data. An attempt also was made to hold "set size" constant. The results indicated that recall was highest with strong cues, intermediate with weak cues, and lowest when no cues were provided. This relative ordering was obtained for both rhyme and synonym cues, suggesting that when a priori strength of relationship has been equated, at either weak or strong levels, both types of cues are equivalent as aids for retrieval. Implications for the encoding specificity principle and for generation-recognition models of free recall were discussed.

Words are represented in memory as complexes of distinctive features. This conjecture represents the principal assumption of several recent conceptualizations of word coding processes (e.g., Anderson & Bower, 1972; Nelson, 1972; Tulving & Thompson, 1973). These features are empirically identifiable as graphic, phonetic, associative, semantic, and imaginal, but they could plausibly consist of any discriminable attributes. According to one version of the multifeature hypothesis, the process of coding a word that is presumably already within *S*'s vocabulary consists of priming one or more of its distinctive features during the study portion of the current learning task (Nelson, 1972). Thus, when a word is presented along with an instruction to remember it, one or more of its attributes are either selectively or automatically primed, i.e., they become functionally effective in the task. Another assumption of this conception is that the priming of specific attributes can be influenced by task characteristics, contextual cues, and instructional sets, as well as by item attributes themselves, and that these features, once effective, function as redintegrative cues for subsequent stimulus retrieval.

One implication of this conceptualization is that items that cannot be recalled under noncued conditions might be made accessible in the presence of extralist cues that prime features shared by specific list items (cf. Tulving & Pearlstone, 1966). Strong extralist associates of list items (e.g., Bahrick, 1969), names of conceptual categories (e.g., Tulving & Psotka, 1971), hononyms and synonyms (Light, 1972) as well as rhymes and graphic cues (Bregman, 1968; Nelson & Brooks, 1974), have all been used as effective recall aids. However, with some exceptions (e.g., Bregman, 1968; Light, 1972), very little effort in this area has been directed toward assessing the relative efficiency of various types of cues. Although the apparent differential emphasis given semantic factors by psychological theorists leaves the impression that meaning is the sine qua non of memory, there is ample evidence suggesting that sensory features, both graphic and phonetic, play a prominant role in word processing (e.g., Gibson, 1971; Nelson, 1972). These considerations suggest that comparisons of the effectiveness of meaning and sensory features as extralist retrieval aids might shed some light on the relative importance of these attributes within the memory representation of a given word. Therefore, this

[1] This research was supported by Grant MH-16360 to the first author from the National Institute of Mental Health and by Stipend GZ-2341 from the National Science Foundation to the second author. The authors are indebted to Eileen Edmunson and Kris Emden for their aid in collecting the data for this experiment.

[2] Requests for reprints should be sent to Douglas L. Nelson, Department of Psychology, University of South Florida, Tampa, Florida 33620.

experiment directly compared rhymes and synonyms as retrieval cues for rendering access to words stored during a rapidly paced single presentation. In order to ensure that this comparison was on a sound basis, two potential problems had to be resolved. First, the a priori similarity between the synonym cues and their targets had to be equated to the a priori similarity between the rhyme cues and their targets. If not so equated, one type of cue may prove to be superior to or not different from the other because of an initial confounding of the cue-to-target relationship. Second, "set size" had to be equalized for both types of cues. If cued recall is analogous to a search task, then the cue that provides the greatest reduction of the number of relevant items will be the most efficient. Providing that the procedures used to eliminate these potential confoundings are effective, equivalent facilitation would suggest that the two types of cues were equally effective, while differential facilitation would imply that one type of cue provided superior retrieval access relative to the other. This experiment, therefore, was explicitly designed to manipulate type of cue—synonym or rhyme—at two levels of cue-to-target relationship, while, at the same time, the a priori relations between cue and target and set size were carefully equated for both types of cues.

Method

Design. The experimental design was a 2^5 between-within factorial. Type of cue, type of recall, lists, and *E*s were all between-*S*s factors and strength relation of cue to target was a within-*S*s factor. Cues were rhymes or synonyms, recall was cued or free, and strength was either strong or weak. Separate free-recall groups were used for rhyme and synonym conditions, since the targets for each of these lists had to be composed of different sets of words.

Materials. All words used in this experiment were chosen from norms obtained from a separate intact group of 74 *S*s in an ongoing psychology course. Each *S* was given a booklet containing 6 pages. The initial page contained rhyme instructions for half of the *S*s and synonym instructions for the remaining half of the *S*s, and the next 5 pages each contained 1 column of 20 words with a blank to the right of each item. The rhyme instructions indicated that on the blank beside each word they should write the *first* word that they could think of that *rhymes* with the word on the left and, similarly, the synonym instructions indicated that they should write down the *first* word that they could think of that means the same thing as the word on the left. The same 100 words were used with both instructional sets, and these words had been chosen by *E* in advance so that each item had several rhymes and synonyms. The order of the pages containing the sample was independently randomized for each *S*.

The normative data provided by these *S*s was then used to construct 4 lists, 2 for the rhyme condition and 2 for the synonym condition. Each list contained 24 pairs, the cue words having served as stimuli in the controlled association task and the targets having been provided by various proportions of *S*s in that task. For half of the pairs, the cue words were strongly associated to their respective targets. An average of 52% of the *S*s provided the target in response to the cue word and, for all of these pairs, the target was the primary response to the cue. For the remaining half of the pairs within a given list, the cue-to-target relation was weak, with an average of 7% of the *S*s responding with that particular target. The average numbers of responses given more frequently than response words selected as targets were 2.83 and 2.63, respectively, for synonyms and rhymes. Furthermore, of all the responses given during the recall phase of the experiment, 22.1% were given more frequently in the controlled association task than the targets used in the weak-synonym-cue condition, and 22.7% were given more frequently than the targets used in the weak-rhyme-cue condition. Thus, strength of relationship was carefully equated at each level for rhyme-related and synonym-related pairs. In addition, to preclude the possibility that the effects of type or strength of cue might be attributed to confoundings associated with differences in [illegible], an attempt was made to equate this variable in all conditions. Potential set size was estimated from the number of different responses given to each cue in the controlled association task. Thus, for the rhyme-related and synonym-related pairs there were an average of 6.83 ($SD = 1.74$) and 6.42 ($SD = 2.36$), respectively, different targets provided for each cue. To control for potential set size effects within the strength variable, the cues having a strong relation to their respective targets in one list had a weak relation to their targets in the other list. For example, in the rhyme condition, KEG was used as a strong cue for the target LEG in one of the lists and as a weak cue for the target PEG in the second list. In the synonym condition, FABLE was used as a strong cue for the target STORY in one of the lists and as a weak cue for the target LEGEND in its complementary list. This procedure resulted in using the same cues for each of the 2 lists in the rhyme condition and for each of the 2 lists in the synonym condition. The targets were, however, different for all 4 lists and, although it was possible to equate word frequency within types of cues, it was impossible to equate relative frequency within

the strength variable. Thus, for each of the 2 lists in the rhyme and synonym conditions, the targets having a strong relation to their cues had an average frequency of occurrence of approximately 80 times per million (Kučera & Francis, 1967). The targets having a weak relation to their cues had an average frequency of occurrence of approximately 17 times per million. This confounding of strength and frequency of occurrence was unavoidable since, in the controlled association task, words provided by the smallest percentage of the *Ss* were also less frequent than words provided by the largest percentage of the *Ss*. Because of this confounding, separate free-recall control groups were used for each list of targets. In addition to providing a direct assessment of the frequency variable, this procedure permitted evaluations of the effectiveness of strong and weak rhyme and synonym cues relative to a free-recall baseline appropriate for each frequency level.

Procedure. All *Ss* participated in individual sessions. On study trials, each list word was shown by a Kodak Carousel slide protector driven by an auxiliary timer at a 2-sec. rate. All words were typed in uppercase. The *Ss* were instructed to pronounce the items aloud as each one appeared and, since the list would be shown only once, to concentrate in order to remember as many items as possible. Immediately following the presentation of the last word, instructions for recall were read. In the cued-recall groups, these instructions specified the nature of the relationship of the cue words to the targets and indicated that using the cues would aid their recall. Thus, the information concerning the nature of the retrieval cues was provided *after* list presentation. In the free-recall control groups, these instructions indicated that a row of asterisks would be repeatedly flashed on the screen, and that each time the asterisks appeared, a word—any word from the list—should be recalled. Both sets of instructions were carefully equated for length. In all conditions, recall was oral and was response paced, i.e., as soon as the target was recalled or *S* indicated that he could not remember the appropriate word, the next cue or row of asterisks was shown. The order of presentation of the targets on the study trial and the order of presentation of the cues on the test trial, if any were independently randomized for every *S*.

TABLE 1

Mean Number of List Words Correctly Recalled

Strength of cue	Rhyme: Cued recall	Rhyme: Free recall	Synonym: Cued recall	Synonym: Free recall
Weak	9.20	4.70	9.00	4.50
Strong	6.65	4.75	5.80	4.35

Subjects. Ten *Ss* were assigned to each list within each condition, and each *E* collected data from 5 of these *Ss*. This procedure resulted in 20 *Ss* within each of the major conditions, and a total of 80 in the entire experiment. For each *E*, *Ss* were assigned to conditions in blocks of 8, with 1 *S* from each condition and list per block. Assignment within blocks was determined by a table of random numbers. All *Ss* were selected from introductory psychology courses and received points toward their grades for participation.

Results

Table 1 presents the mean number of correct recalls as a function of type of cue, type of recall, and strength of relation of cue to target. An analysis of variance of these data indicated that type of recall, $F(1, 76) = 94.61, p < .01$; strength of cue, $F(1, 76) = 22.93, p < .01$; and the Type of Recall × Strength of Cue interaction, $F(1, 76) = 21.40, p < .01$, were the only significant sources of variance. Type of cue and all other interactions were not reliable. When recall was free, an average of 4.60 items were recalled in the strong-cue (high frequency) condition, and an average of 4.55 items were recalled in the weak-cue (low frequency) condition. When recall was cued, an average of 9.10 items were recalled when cues were strong, and an average of 6.22 items were recalled when cues were weak. Fisher's least significant difference for this interaction was .87. Thus, when recall was free, there were no differences between the strong- and weak-cue conditions, i.e., word frequency had no reliable effects. When recall was cued, significantly more items were recalled when strong as compared to when weak cues were provided. Furthermore, comparison of recall performance in the weak-cue condition to the free-recall conditions indicated that weak cues significantly facilitated recall compared with no cues at all. Thus, recall was highest when strong cues were provided, intermediate when weak cues were provided, and poorest when no retrieval aids were presented. The facilitating effects obtained with weak cues contrast with the inhibiting effects obtained with comparable cues used by Thompson and Tulving (1970).

Discussion

When set size and a priori strength of relation of cue to target have been equated, synonyms and rhymes are apparently equally effective as extralist retrieval aids. This equivalency is obtained at both strong and weak levels of associative strength. Therefore, cuing with either semantic or sensory attributes can provide equally effective access to the coded representations of target words primed in the context of a rapidly presented list of unrelated items. Presumably, the two types of cues are effective either because they provide independent access routes to two different representations of the target word, one based on sensory attributes and one based on semantic attributes, or because they provide access to a multifeatured representation that contains both types of information. Irrespective of the correctness of either of these alternatives, the primary implication of the results of this experiment is that the sensory attributes of a word may be as functionally important in its representation as are its semantic attributes.

These results also carry implications for two theories concerning extralist cuing effects, the encoding specificity principle (Tulving & Thompson, 1973) and the generation–recognition models (Anderson & Bower, 1972; Bahrick, 1970). According to the encoding specificity principle, the effectiveness of an extralist cue is completely determined by what is encoded while the target items are being studied. A retrieval cue cannot be effective unless the target is specifically encoded with respect to that cue at the time of its storage. The generation–recognition models assume that the extralist cue serves to generate relevant response alternatives and that recognition tests are applied to each item. Items meeting the criteria of acceptability are produced as overt responses. These models make no explicit assumptions concerning the relation between the effectiveness of the extralist cue and the manner of storage of the target item. However, most of them assume that words are represented in memory as lists or collections of features that specify their sound, meaning, and other attributes and, presumably, most might easily incorporate the notion that the representation of a word might be modified when presented with a context word that primes a specific attribute. Thus, at least in general orientation, the theories are similar in many respects.

To account for the facilitating effects of cuing in this study, the encoding specificity principle would have to assume that both semantic and sensory features of the target words were primed during the encoding phase of the task. False recognitions obtained by using synonym distractors in a recognition memory paradigm attest to the reasonableness of the assumption that meaning attributes are primed during the presentation of a list of unrelated words (e.g., Anisfeld & Knapp, 1968). Furthermore, false recognition and transfer data using homophones are consistent with the assumption that sensory attributes also are primed in this situation (Nelson & Davis, 1972), and therefore, extralist cuing effects obtained with these stimuli (Light, 1972) can be interpreted as being compatible with this principle. Thus, synonyms and rhymes were presumably effective cues because they provided sufficient redintegrative information about their respective targets. However, although this interpretation accounts for the cuing effects per se, it fails to explain why differences were obtained between strong and weak rhyme cues. Words at either level provide the same degree of overlap of sensory features and, therefore, in contrast to synonym cues, there should have been no strength effect associated with these cues. This strength effect might be explained by the assumption that, given the cue, *S*s simply guessed the first rhyme that came to mind. This possibility would account for some portion of the cuing effect associated with strong [illegible]. (Bahrick, 1969). However, simple guessing would appear inadequate as an explanation for the cuing effect associated with weak cues since, if *S*s were saying the first word that occurred to them, nearly all of their responses would have been intrusions. Alternatively, the strength effect might be explained by assuming that, in cued recall, the retrieval aid acts to initiate a search through items sharing features with the cue word and, as items are generated, recognition checks are performed (Anderson & Bower, 1972; Bahrick, 1970). If it can be assumed that words associated with each type of cue are generated hierarchically, then the chances of producing a list word are greater when extralist cues serve as cues for responses that are high in the hierarchy (strong cues) than for those low in the hierarchy (weak cues). This explanation for the facilitating effects of cuing incorporates the major assumptions of both the encoding specificity hypothesis and the generation–recognition models.

REFERENCES

Anderson, D. R., & Bower, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, **79**, 97–123.

Anisfeld, M., & Knapp, M. Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology*, 1968, **77**, 171–179.

Bahrick, H. P. Measurement of memory by prompted recall. *Journal of Experimental Psychology*, 1969, **79**, 213–219.

Bahrick, H. P. A two-phase model for prompted recall. *Psychological Review*, 1970, **77**, 215–222.

Bregman, A. S. Forgetting curves with semantic, phonetic, graphic, and contiguity cues. *Journal of Experimental Psychology*, 1968, **78**, 539–546.

Gibson, E. J. Perceptual learning and the theory of word perception. *Cognitive Psychology*, 1971, **2**, 351–368.

Kučera, H., & Francis, W. N. *Computational analysis of present-day American English.* Providence: Brown University Press, 1967.

Light, L. L. Homonyms and synonyms at retrieval cues. *Journal of Experimental Psychology*, 1972, **96**, 255–262.

Nelson, D. L. Words as sets of features: The role of phonological attributes. In R. F. Thompson & J. F. Voss (Eds.), *Topics in learning and performance.* New York: Academic Press, 1972.

Nelson, D. L., & Brooks, D. H. Retroactive inhibition of rhyme categories in free recall: Inaccessibility and unavailability of information. *Journal of Experimental Psychology*, 1974, **102**, 277–283.

Nelson, D. L., & Davis, M. J. Transfer and false recognitions based on phonetic identities of words. *Journal of Experimental Psychology*, 1972, **92**, 347–353.

Shapiro, S. I., & Palermo, D. S. An atlas of normative free association data. *Psychonomic Monograph Supplement*, 1965, **2** (Whole No. 28).

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

Tulving, E., & Psotka, J. Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology*, 1971, **87**, 1–8.

Tulving, E., & Thompson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352–373.

(Received July 9, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 508–511

RHYTHM IS PROCESSED BY THE SPEECH HEMISPHERE[1]

GEORGE M. ROBINSON[2] AND DEBORAH JO SOLOMON

Duke University

Following Lashley and Neisser, models of speech cognition frequently stress a rhythmic organizing principle and would predict that speech processing is intimately related to the processing of rhythmic patterns. However, in agreement with clinical data, dichotic listening studies establish that, while speech stimuli are processed by the left hemisphere, other nonspeech auditory stimuli are processed by the right. Challenging this distinction, the present study shows that nonspeech rhythmic patterns carrying no phonetic information are processed by the same hemisphere as speech. Twenty-four *Ss* listened to 30 dichotic pairs of rhythmic pure-tone patterns. In a counterbalanced, forced-choice recognition task, they were able to correctly identify the pattern heard in the right ear significantly more often than the left. The results suggest that since rhythmic patterns are the only nonspeech auditory stimuli to share the processing of the left hemisphere with speech, models involving rhythmic organization in speech cognition are to be encouraged. Since both speech and rhythm require hierarchical organization, it is likely that the left hemisphere is better able to process hierarchically.

Lashley's (1951) classic paper focused attention on the problems of organization in speech cognition. He proposed that the temporal order of the elements of speech results from an underlying hierarchical structure. The regular appearance of a series of elements with respect to time has recently been described as a rhythmic phenomenon.

> One way to think of the effect of rhythm is that it may provide a set of reference points to which digits and words can be attached. . . . A rhythmic pattern is a structure which serves as a support, an integrator and a series of cues for the words to be remembered [Neisser, 1967, pp. 223, 235].

In a recent review, Martin (1972) presented "a good deal of evidence that rhythmic patterning carries a heavy information load in ordinary connected speech [p. 500]."

In the past decade, several converging lines of research have supported and refined the classical notion that language functions are primarily mediated by the left cerebral hemisphere. Such hemispheric specialization is supported by studies on patients with unilateral lesions and excisions (e.g., Milner, 1971) and cerebral commisurotomy (e.g., Milner, Taylor, & Sperry, 1968). For normal *Ss*, the asymmetries found using the dichotic listening technique, in which different stimuli are presented simultaneously, one to each ear, have closely paralleled the clinical studies. Speech material entering the right ear is processed more effectively, while the opposite is true for nonspeech material (Kimura, 1973; Studdert-Kennedy & Shankweiler, 1970). The resulting conclusion that contralateral auditory pathways are stronger than ipsilateral ones is independently supported by electrophysiological evidence (e.g., Rosenzweig, 1951). Thus, the dichotic listening technique has become a tool for investigating hemispheric specialization.

Given the well-established laterality of speech functions, theoretical models that involve rhythmic organization in speech processing would predict that the hemisphere that is more effective in processing speech will also be the more effective hemisphere for rhythm. Since a wide variety of nonspeech auditory stimuli, including music, were demonstrated to be processed better by the right hemisphere (Dee, 1971; Gordon, 1970; Kimura, 1964;

[1] Thank you Charles W. White and John S. [illegible] Health Service Grant 1-R03-MH21981-01 supported the research.

[2] Requests for reprints should be sent to George M. Robinson, Department of Psychology, Duke University, Durham, North Carolina 27706.

Milner, 1962; Spellacy, 1970), such a prediction is far from obvious. In fact, both Milner and Spellacy, expecting the right hemisphere to be superior, tested for laterality of rhythm but found no significant effect.

Halperin, Nachshon, and Carmon (1973) pointed out, however, that the speech vs. nonspeech distinction is no longer adequate to describe hemispheric specialization for auditory stimuli. Their *Ss* were able to report the order of 3 broad-band noises heard in the right ear more accurately than in the left when the frequency and bandwidth or pulse durations included 1 or 2 changes. For noises without these changes, the left ear was superior. A related finding is the one by Carmon and Nachshon (1971) that left-hemisphere lesions impair the ability to correctly report the order of a series of sounds and colored lights. Halperin et al. concluded that these findings suggest

> that the left hemisphere specializes in the processing of temporal patterns of stimuli and that perception of verbal stimuli might be associated with sequential analysis of acoustic stimuli [p. 49].

Some cautions should be observed in interpreting the results of Halperin et al. (1973). The frequency shifts and bandwidth changes in the noises used in their first study may have directly involved the phonetic analyzing ability of the speech hemisphere. Thus, instead of attributing the increasing dominance of the right ear to increasing complexity of the temporal pattern, the shift may have occurred because changes in the spectrum were actually encoded in a way similar to speech encoding. Their second experiment, using changes in the durations of 3 noise bursts, comes much closer to isolating temporal complexity as the factor responsible for shifting dominance from the right to the left hemisphere. It may be, however, that the abrupt onset and offset of the broad-band noises were not so unlike speech sounds. In fact, their data do not show that the right hemisphere was dominant when there were zero duration changes.

The following experiment attempted to directly investigate the hypothesis that complex nonspeech temporal patterns containing no phonetic information (rhythms) are processed by the speech hemisphere.

Method

Subjects. Twenty-four Duke University students with normal hearing were unobtrusively screened to eliminate left-handers by noting which hand they used to sign their names. After the listening task, *Ss* were further examined for handedness. None were ambiguous cases. The *Ss* were volunteers without any knowledge of the purpose of the study. None were familiar with Morse code.

Stimuli and procedure. Each rhythm consisted of a "refrain" repeated 3 times with a 50-msec. interval. Each refrain consisted of a random series of 4 7 short (50-msec.) and long (150-msec.) 900-Hz. sinusoidal pulses separated by 50 msec. The durations were chosen so that both the short and long pulses with their interpulse intervals would fall well within the range of spoken syllable durations. Each of the 30 trials consisted of a dichotic pair of rhythms followed by a random series of 4 individual recognition test rhythms, binaurally presented: the 2 components of the dichotic pair and 2 new ones. These 4 recognition test rhythms were separated from the dichotic rhythm and from each other by 4 sec., with a 10-sec. intertrial interval. Within each trial, all 6 rhythms had identical durations. The components of the dichotic pair began and ended simultaneously.

The test materials were generated by a PDP-8I. computer, recorded on a Revox A77 stereo tape recorder and played through Koss PRO-4A earphones at 60-db. SPL.

Each *S* first balanced the 2 channels for a binaural series of regular pulses so that the sound appeared to be centered in his head. The *Ss* then listened to the test materials in 1 of 4 conditions: counterbalanced by interchanging left and right channels of the tape and interchanging the 2 half blocks of 15 trials. The *Ss* were required in a forced-choice paradigm to mark 2 of a row of 4 boxes to indicate which 2 of the 4 recognition choices they believed to be part of the dichotic pair. No letters or numerals appeared anywhere on the coding sheets.

Results

Table 1 shows the number and percentage of correct identifications out of 30 trials for the right and left ears of each *S*. The right-ear scores were significantly better: $t = 3.01$, $p < .01$, 2-tailed test. The scores for each ear were both better than chance: for the right ear, $t = 6.45$, $p < .001$, and for the left ear, $t = 2.87$, $p < .01$, 2-tailed tests.

Discussion

The results show that rhythmic patterns unlike other nonspeech auditory stimuli, are

TABLE 1

Number and Percentage of Correct Responses in 30 Trials for 24 *Ss*

S	Right ear		Left ear	
	No. correct	Percentage	No. correct	Percentage
1	19	63	13	43
2	21	70	14	47
3	21	70	19	63
4	17	57	20	67
5	25	83	16	53
6	22	73	20	67
7	20	67	18	60
8	15	50	16	53
9	20	67	13	43
10	21	70	12	40
11	17	57	21	70
12	18	60	17	57
13	21	70	15	50
14	17	57	15	50
15	16	53	16	53
16	17	57	21	70
17	17	57	19	63
18	22	73	17	57
19	17	57	16	53
20	15	50	12	40
21	13	43	18	60
22	18	60	17	57
23	20	67	16	53
24	20	67	16	53
$\bar{X}$	18.71	62	16.54	55
SD	2.76		2.58	

processed better by the same hemisphere that is dominant for speech stimuli. This study supports the heuristic discussions of Lashley (1951) and Neisser (1967). The results confirm the suggestion by Halperin et al. (1973) that temporal patterns as well as speech stimuli are processed by the left hemisphere. Because rhythm is the only nonspeech auditory feature found to be processed by the speech hemisphere, models of speech cognition based on rhythmic organization are to be encouraged.

Because the rhythms used in the present experiment were moderately complex, rapid, and repeated 3 times, it is likely that they were processed hierarchically in the manner described by Lashley (1951) and Neisser (1967). The interested reader will find a discussion of how the details of the hierarchic structures underlying rhythmic patterns might be organized in Martin (1972). It is probable that the left hemisphere is dominant for both nonspeech rhythms and speech stimuli because it is better able to do the hierarchical processing they both require.

The dichotic listening technique demonstrates directly only a relationship between the performance of one ear and handedness. Two inferences permit conclusions about whether or not a particular stimulus is processed better by the speech hemisphere. The first was discussed in the introduction, that is, that contralateral auditory connections are stronger than ipsilateral ones. In this study, material entering the right ear was processed (mainly) by the left hemisphere. The second inference is that right-handed *Ss* do their speech processing with the left hemisphere. Branch, Milner, and Rasmussen (1964), by comparing the effect on speech of injecting sodium amytal into first one and then the other of their *Ss*' carotid arteries, determined that 90% of right-handers and 60% of left-handers have their speech functions mediated by the left hemisphere.

On the assumption that the hypothesis is correct, we should not be surprised that our data show (as do other dichotic studies) a few *Ss* to have higher scores for their left ears. Although it is possible that this is a major source of variance, there is no independent means to assess it within this study. While the distribution tends to be bimodal, the tendency is not significant.

Halperin et al. (1973) asked their *Ss* to name the noises in the order they appeared. In our study, the coding sheet was deliberately kept free of language-related material to avoid biasing the data by "priming" the left hemisphere. The spatial array used would, if anything, invite right-hemisphere processing. In addition, the recognition test rhythms for the left and right ears appeared in random order. These procedures avoid the bias claimed by theories which attribute the dichotic listening effect to differences in the order of recall, claiming a greater short-term memory loss for material played to the left ear (e.g., Inglis, 1965). Such caution may not be necessary. Bryden (1967) reviewed the several explanations for dichotic listening asymmetry and concluded that Kimura's (1961) theory attributing the asymmetry to perceptual differences is the only one that is consistent with all experimental results.

REFERENCES

Branch, C., Milner, B., & Rasmussen, T. Intracarotid sodium amytal for the lateralization of cerebral speech dominance. *Journal of Neurosurgery*, 1964, 21, 399–405.

Bryden, M. P. An evaluation of some models of

dichotic listening. *Acta Oto-Laryngologica*, 1967, **63**, 595–604.

Carmon, A., & Nachshon, I. Effect of unilateral brain damage on perception of temporal order. *Cortex*, 1971, **7**, 410–418.

Dee, H. L. Auditory asymmetry and strength of manual preference. *Cortex*, 1971, **7**, 236–245.

Gordon, H. W. Hemispheric asymmetries in the perception of musical chords. *Cortex*, 1970, **6**, 387–398.

Halperin, Y., Nachshon, I., & Carmon, A. Shift of ear superiority in dichotic listening to temporally patterned nonverbal stimuli. *Journal of the Acoustical Society of America*, 1973, **53**, 46–50.

Inglis, J. Dichotic listening and cerebral dominance. *Acta Oto-Laryngologica*, 1965, **60**, 231–238.

Kimura, D. Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, 1961, **15**, 166–171.

Kimura, D. Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 355–358.

Kimura, D. The asymmetry of the human brain. *Scientific American*, 1973, **228** (12), 70–78.

Lashley, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley, 1951.

Martin, J. G. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 1972, **79**, 487–509.

Milner, B. Interhemispheric differences and psychological processes. *British Medical Bulletin*, 1971, **27**, 272–277.

Milner, B. Laterality effects in audition. In V. B. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance*. Baltimore: Johns Hopkins Press, 1962.

Milner, B., Taylor, L., & Sperry, R. W. Lateralized suppression of dichotically presented digits after commisural section in man. *Science*, 1968, **161**, 184–186.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Rosenzweig, M. R. Representation of the two ears at the auditory cortex. *American Journal of Physiology*, 1951, **167**, 147–158.

Spellacy, F. Lateral preferences in the identification of patterned stimuli. *Journal of the Acoustical Society of America*, 1970, **47**, 574–577.

Studdert-Kennedy, M., & Shankweiler, D. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, **48**, 579–594.

(Received July 9, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 512-516

PERFORMANCE DURING CONTINUOUS AND INTERMITTENT NOISE AND WEARING EAR PROTECTION

L. R. HARTLEY[1]

Medical Research Council, Applied Psychology Unit, Cambridge, England

The *S*s were tested on the 5-Choice Test of Serial Reaction for 40 min. in intermittent and continuous broad band noise (Experiment I) and while wearing ear protection in continuous noise (Experiment II). Gaps in performance during intermittent noise were approximately half those in continuous noise. Errors were affected equally adversely by both intermittent and continuous noise. Ear protection interacted with noise and time-on-task, reducing gaps in noise in the first half but not the second half of the test. Results are consistent with the theory that there are 2 different effects of noise. The improvement of performance when noise was intermittent was attributed to a reduction in the monotony experienced during a long exposure to continuous noise. The value of ear protection was attributed to a reduction in perceived loudness and prevention of temporary arousal following the onset of noise.

One recent interpretation of the adverse effect of noise was that it functioned as a distractor (Broadbent, 1958) causing impairment in tasks that had characteristics rendering them sensitive to distractions. However, evidence against this view arose from the beneficial effect of noise on performance of the 5-Choice Test of Serial Reaction following sleep loss (Corcoran, 1962; Wilkinson, 1963) when both conditions, applied separately, caused impairment. In addition, incentive augmented the adverse effect of noise as compared to its normal beneficial action (Wilkinson, 1963). These results suggested that the adverse effect of noise was mediated by an increase in arousal which offset the effect of sleeplessness but which was further enhanced by incentive. One difficulty with this view was that the arousing effect of noise, and consequent change in performance, would probably be greatest when the noise was first introduced, but the observed adverse effect generally increased with duration of noise exposure, independent of time-on-task (Hartley, 1973). The results of Hartley's experiment, which also involved the 5-Choice Test of Serial Reaction, ruled out the possibility that the adverse effect of noise was dependent on prior performance of the test, since noise introduced after many minutes of performance had virtually no effect. Consequently, the appearance of the adverse effect of noise after exposure was attributed to an accumulating function of the noise.

This conclusion was substantiated by performance on a Stroop color interference test in noise (Hartley & Adams, 1974). Interference was increased by 20-min. prior exposure to noise as compared to quiet. An additional finding of interest in Hartley and Adams' experiment (also, Houston, 1969; Houston & Jones, 1967; O'Malley & Poplawsky, 1971) was that interference was reduced if noise was given only during performance and not also prior to it, suggesting different effects of short and long noise exposures. There is some evidence that reduced interference accompanies a raised level of arousal. Agnew and Agnew (1963) and Tecce and Happ (1964), using threat of electric shock, and Callaway (1959), using amphetamine, obtained results indicating reduced interference under these conditions; a short exposure to noise may thus be arousing. The explanation of the increase in interference following a long exposure could be in similar terms. The increase

[1] Requests for reprints should be sent to L. R. Hartley, Medical Research Council, Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, England.

Now at University of Otago, Dunedin, New Zealand.

TABLE 1

MEAN NUMBER OF GAPS IN EACH 10-MIN. BLOCK UNDER EACH CONDITION

Condition	Knowledge of results					No knowledge of results				
	Block 1	Block 2	Block 3	Block 4	Total	Block 1	Block 2	Block 3	Block 4	Total
Q	1.44	2.67	3.78	4.00	11.89	1.83	6.11	5.44	6.00	19.38
IN	4.78	5.39	6.83	7.44	24.44	4.56	6.44	8.06	8.17	27.23
CN	4.39	8.28	9.78	9.50	31.95	6.33	10.17	11.67	10.94	39.11

Note. Abbreviations: Q = quiet, CN = continuous noise, and IN = intermittent noise.

may indicate some form of reduced arousal, which would be consistent with results indicating increased interference following amylobarbital (Callaway, 1959) and hyoscine (Ostfeld & Arugete, 1962). The initial arousing effect of noise could be attributed to the onset of the intense stimulation to which *S* later adapted. The later reduction in arousal could be attributed to the monotony and sensory isolation accompanying the longer exposure to noise.

To test these suggestions, 2 experiments were considered. If the later part of the adverse effect of continuous noise is due to the monotony and lack of variability, then presenting the noise intermittently should reduce this part of the adverse effect. Secondly, if the adverse reaction following the onset of noise is due to its loudness or intensity, then wearing ear protection should be beneficial at the start of the exposure. Wearing ear protection should not be of any further benefit when adaptation to loudness has occurred and the adverse effect of monotony predominates.

EXPERIMENT I: INTERMITTENT AND CONTINUOUS NOISE

Performance of *S*s on the 5-Choice Test of Serial Reaction was compared during quiet (Q), intermittent noise (IN), and continuous noise (CN) conditions.

Method

Subjects. There were 2 groups of *S*s. Each group consisted of 18 housewives and professional men, ranging 18–45 yr. of age and totaling 36 *S*s in all. All possessed normal hearing (i.e., an absence of hearing loss greater than 35 db. in one ear or 30 db. in both ears at any of 7 test frequencies covering the audible range).

Apparatus. Performance was measured with Leonard's (1959) 5-choice serial reaction apparatus, which requires that *S* be seated before a display consisting of 5 neon light sources arranged in a pentagon, one of which is always illuminated. Arranged on a horizontal response board are 5 brass disks corresponding to the light sources. The *S* is required to tap the disk appropriate to the lamp illuminated. The light promptly extinguishes and another is lit. The *S* works as quickly and accurately as possible, scoring as many corrects and as few errors as possible. The third score consists of pauses or gaps of 1½ sec. between tapping one disk and tapping the next. The *S*s in this study performed with and without knowledge of their results (KR and NKR). Immediate knowledge of gaps and errors was furnished by a brief illumination of a lamp on the left of the response board if an error was made, and by a brief illumination of a lamp on the right of the board if a gap was made. Knowledge of the number of correct responses made in each test was provided at the end of each test.

Procedure. Following 40-min. preliminary practice on the test, each *S* performed the test for 40 min. in Q at 70 db., CN at 95 db., and IN at 70 and 95 db., measured on the C scale of the sound level meter. Although *S*s were tested at different times of day, each *S* was given 1 treatment per day at the same time of day. The order of presentation of the 3 treatments was determined by counterbalanced Latin squares for each group of 18 *S*s. One group of 18 *S*s performed the test with KR and the other group performed with NKR. The former group received brief flashes from the indicator lights at the side of the response board whenever a gap or error was made, and they were also shown the total number of gaps, corrects, and errors made by each *S* in each test in a chart. Broad-band noise having equal energy per octave was used in all conditions. The IN condition consisted of alternate presentations of 95 and 70 db. The mean duration of the 70-db. interval was 1.5 sec., with a range .3–2.7 sec. The mean duration of the 95-db. interval was 3 sec., with a range .6–5.4 sec.

Results

Performance was scored in 10-min. blocks for each 40-min. test. Table 1

TABLE 2

MEAN NUMBER OF ERRORS IN EACH 10-MIN. BLOCK UNDER EACH [illegible]

Condition	Knowledge of results					No knowl[illegible]				
	Block 1	Block 2	Block 3	Block 4	Total	Block 1	Block 2	Block 3	Block 4	Total
Q	3.17	4.94	4.17	3.61	15.89	3.72	4.11	5.44	6.72	19.99
IN	5.11	5.83	4.22	4.89	20.05	5.39	7.72	10.00	8.50	31.61
CN	3.33	6.50	6.39	5.89	22.11	6.89	5.94	7.17	9.44	29.44

Note. Abbreviations: Q = quiet, CN = continuous noise, and IN = intermittent noise.

shows mean gaps and Table 2 shows mean errors for KR and NKR for each experimental condition. These scores were analyzed using analysis of variance as described by Winer (1962, p. 329) following arc sine square root transformation of the data. Considering gaps, there was no difference between the KR and NKR groups, F (1, 34) = .86, $p > .10$. Knowledge of results did not interact with any of the other treatments, and hence results from the 2 groups were pooled together for further analysis. There was a large adverse effect of the noise conditions upon the number of gaps made, F (2, 68) = 21.92, $p < .001$. Mean number of gaps in Q was 15.6, and in IN, 25.8. This increase ($p < .01$) was very similar to the increase in gaps in CN ($\bar{X}$ gaps = 35.5) as compared to IN ($p < .01$). The noise conditions did not interact with any other factor.

Considering errors, there was no overall reliable difference between KR and NKR, but more errors were made in the last 10 min. of the test with no knowledge than with knowledge, F (1, 272) = 3.91, $p < .05$. Pooling the NR and NKR conditions together, the noise conditions had a reliable effect upon errors, F (2, 68) = 4.49, $p < .025$. Compared to Q, the mean increase of 7.89 errors in IN ($p < 0.05$) was virtually identical to the mean increase of 7.84 errors in CN ($p < .05$). These results indicate that IN caused approximately half of the impairment apparent in the CN condition in terms of gaps, but that the effect of the 2 conditions upon errors was highly similar.

EXPERIMENT II: EAR PROTECTION

The *S*s were tested on the 5-Choice Test of Serial Reaction in Q and noise (N), both with and without wearing ear protection.

Method

Subjects. The *S*s were 16 housewives and professional men, ranging 18–45 yr. of age. All possessed normal hearing (i.e., an absence of hearing loss greater than 35-db. in one ear or 30 db. in both ears at any of 7 test frequencies covering the audible range).

Procedure. Following 40-min. preliminary practice on the 5-Choice Test of Serial Reaction, each *S* performed the test for 40 min. in continuous Q and continuous N, both with and without ear protection. The order of presentation of the 4 treatments was determined by counterbalanced Latin squares (Williams, 1949). Broad-band noise having equal energy per octave was used throughout. In the N conditions it was presented at 95 db., and in the Q conditions, at 70 db., measured on the C scale. Each *S* was given one treatment per day at the same time of day, although *S*s were not all tested at the same time of day.

The ear protection was provided by Amplivox Sonogard ear defenders, specified as providing a substantially linear attentuation with frequency over the audible range (20 db. at 200 Hz., 30 db. at 500 Hz., 38 db. at 1 and 2 kHz., and 45 db. at 4 kHz., measured at one-fifth octave bands and the American Standard Method for the Measurement of the Real Ear Attenuation of Ear Protection at Threshold, ASAZ-24.22-1957).

Results

Performance was scored in 10-min. blocks for each 40-min. test. Table 3 shows mean gaps and Table 4 mean errors for each experimental condition. These scores were analyzed using analysis of variance with repeated measures on each factor, following arc sine square root transformation of the data. Error terms in the F ratios utilized interactions with *S*s.

Considering gaps, as Table 3 shows, there was an overall, adverse effect of N

TABLE 3

Mean Gaps in Each 10-Min. Block under Each Condition

Condition	Block 1	Block 2	Block 3	Block 4	Total
Q	2.57	5.13	6.69	7.31	21.70
Q + ED	4.07	6.57	8.25	7.50	26.39
N	5.06	11.12	10.44	8.94	35.56
N + ED	3.57	6.50	9.81	9.88	29.76

Note. Abbreviations: Q = quiet, N = noise, and ED = ear defenders.

as compared to Q, $F\ (1,\ 15) = 10.88$, $p < .01$. Overall, ear defenders had no effect, $F\ (1,\ 15) = .05$, $p > .10$, but ear defenders interacted with Test Blocks × Noise, $F\ (7,\ 105) = 2.12$, $p < .05$. This interaction was mainly due to the fact that in the first 20 min. of the test, when ear defenders were worn, the N as compared to the Q condition had no adverse effect at all. Without ear defenders, N as compared to Q had a large adverse effect ($p < .02$), reliably larger than when wearing ear defenders ($p < .02$). In the last 20 min. of the test, this antagonistic Ear Defenders × Noise interaction disappeared ($p > .10$), and the benefit of wearing ear defenders in N as compared to Q was reliably less than in the first half of the test ($p < .05$).

As Table 4 shows, errors were not significantly adversely affected by either noise or ear defenders; since the variables did not interact, their analysis was not pursued. These results indicate that ear defenders greatly reduce the adverse effect of the first 20 min. of the N condition but have little or no effect after this point in terms of gaps. Clearly there would be no point in looking for changes in errors

TABLE 4

Mean Errors in Each 10-Min. Block Under each Condition

Condition	Block 1	Block 2	Block 3	Block 4	Total
Q	2.10	4.67	6.11	5.72	18.60
Q + ED	3.30	5.22	8.80	6.25	23.57
N	2.87	6.10	8.10	5.19	22.26
N + ED	3.45	9.90	9.73	7.97	31.05

Note. Abbreviations: Q = quiet, N = noise, and ED = ear defenders.

if noise and ear defenders had no reliable effects on this score.

Discussion

In general, the pattern of results from both experiments is consistent with the suggestion of 2 rather different effects on the organism of loud noise. In terms of the introductory hypothesis, IN greatly reduced the adverse effect upon gaps, and although interactions between noise conditions and time-on-task were not apparent, it is clear from the means in Table 1 that the beneficial effect of intermittency developed only after the first 10 min. of exposure. In the first block of a test, IN was approximately similar to CN in its adverse effect, but substantially less in the following part of the test. This pattern fits in well with the expectation that any initial adverse effect would, in the main, be due to arousal, but that any later decrement would be due to monotony which would be partially alleviated by intermittency or variability in the noise.

Still considering gaps, the benefit of wearing ear protection appears to be restricted to the first 20 min. of exposure, suggesting that intensity causes the initial adverse effect, which is consistent with the arousal interpretation. Similarly, the absence of any such benefit in the latter half of the test points to an increasing influence of monotony and isolation on performance, unaffected by ear protection.

Although the changes in gaps support the view that there are 2 different effects of noise, the rather similar effects of IN and CN on errors has to be accounted for. The prevailing view of the relationship between gaps and errors has been that they represent opposite points on the speed and error dimension. Performance under stress has often shown a reciprocal relationship between the 2 scores; noise usually increased errors (Broadbent, 1953) but sleep loss increased gaps (Wilkinson, 1963). Increasing arousal was conceived as moving the criterion in a risky direction and vice versa (Broadbent, 1971). Hartley[2] found just such a reciprocal relationship when noise presented over headphones caused more gaps and fewer errors than noise in the free-field condition. On this view, if the effect of the monotony was to reduce arousal, errors would be expected

[2] L. R. Hartley. Comparison of headphone and free-field noise on the Stroop test. In preparation.

to be more numerous in IN as compared to CN, particularly toward the end of the test when IN should have most benefit in opposing the reduction in arousal during the monotony of the CN condition. There is little evidence here that this is the case. Probably the most satisfactory explanation is that IN and CN are broadly similar in their arousing quality, since they have the same intensity, and that the main benefit of intermittency is in reducing the monotony and isolation of continuous noise rather than increasing the level of arousal. This conclusion would rather imply that monotony does not simply oppose arousal. Indeed there may be 2 mechanisms involved in arousal, one of which is affected by monotony and the other of which is affected by loudness or intensity, in this case. Such a view has support from a number of experiments involving the interaction of stresses, including the combined effect of noise and sleeplessness, alcohol and incentive, and personality variables and stresses (Broadbent, 1971).

REFERENCES

Agnew, N., & Agnew, M. Drive effects on tasks of narrow and broad attention. *Quarterly Journal of Experimental Psychology*, 1963, **15**, 58–62.

Broadbent, D. E. Noise, paced performance and vigilance tasks. *British Journal of Psychology*, 1953, **44**, 295–303.

Broadbent, D. E. *Perception and communication.* London: Pergamon Press, 1958.

Broadbent, D. E. *Decision and stress.* London: Academic Press, 1971.

Callaway, E. The influence of amobarbital (amylobarbitone) and methamphetamine on the focus of attention. *Journal of Mental Science*, 1959, **105**, 382–392.

Corcoran, D. W. J. Noise and loss of sleep. *Quarterly Journal of Experimental Psychology*, 1962, **14**, 178–182.

Hartley, L. R. Effect of prior noise or prior performance on serial reaction. *Journal of Experimental Psychology*, 1973, **101**, 255–261.

Hartley, L. R., & Adams, R. G. Effect of noise on the Stroop test. *Journal of Experimental Psychology*, 1974, **102**, 62–66.

Houston, B. K. Noise, task difficulty, and Stroop color-word performance. *Journal of Experimental Psychology*, 1969, **82**, 403–404.

Houston, B. K., & Jones, T. M. Distraction and Stroop color-word performance. *Journal of Experimental Psychology*, 1967, **74**, 54–56.

Leonard, J. A. *5-Choice serial reaction apparatus.* (Applied Psychology Research Unit Report 326/59.) Cambridge, England: Medical Research Council, 1959.

O'Malley, J. J., & Poplawsky, A. Noise induced arousal and breadth of attention. *Perceptual and Motor Skills*, 1971, **33**, 887–890.

Ostfeld, A. M., & Aruguete, A. Central nervous system effects of hyoscine in man. *Journal of Pharmacology and Experimental Therapeutics*, 1962, **137**, 133–139.

Tecce, J. J., & Happ, S. J. Effects of shock arousal on a card sorting test of colour word interference. *Perceptual and Motor Skills*, 196[illegible], **19**, 905–906.

Wilkinson, R. T. Interaction of noise with knowledge of results and sleep deprivation. *Journal of Experimental Psychology*, 1963, **6**[illegible], 332–337.

Williams, E. J. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research*, 1949, **2**, 149–168.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received July 16, 1973)

Journal of Experimental Psychology
[illegible]

SHORT REPORTS

RECOGNITION AND RECALL AS A FUNCTION OF INSTRUCTIONAL MANIPULATIONS OF ORGANIZATION[1]

ROBERT M. SCHWARTZ[2] AND MICHAEL S. HUMPHREYS[3]

University of British Columbia, Vancouver, British Columbia, Canada

Recognition and recall were examined as a function of organizational instructions. In Experiment I instructions to group items in input order by using visual imagery produced increased recall but did not increase recognition performance. In Experiment II instructions to rehearse overtly the item being presented produced decreased recall but did not decrease recognition performance. The results were discussed as supportive of a 2-process theory of recognition and recall.

A recent controversy concerns the effect of organization on recognition accuracy. Some studies have demonstrated a facilitative effect (e.g., Jacoby, 1972), and other studies have found no effect (e.g., Bruce & Fagan, 1970). One possible reason for the discrepancy is that the manipulations of organization which have been used are manipulations of the structure of the to-be-presented list; that is, *Ss* in experimental and control conditions have not been presented the same lists in the same presentation order, and thus the effects of organization may be confounded with the effects of the nature of the lists. In order to avoid this possible confounding, in the present experiments, the manipulations used involved instructions about processing the list items. With instructional manipulations, *Ss* in experimental and control conditions could be presented the same lists in the same orders. Also, in addition to recognition tests, recall tests were used. The use of recall tests allows differences in organization to be inferred from differences in order of output and/or differences in recall performances.

Experiment I. In Experiment I *Ss* in experimental conditions were instructed to group the words to be recalled in their input order by using visual imagery. The *Ss* in control conditions were given no special instructions. Insofar as organization includes grouping of items at input, the present instructions can be interpreted as manipulations of organization (see, e.g., Bower, Lesgold, & Tieman, 1969).

The stimuli were 120 concrete ($C > 5.50$) and frequent ($F > 20$) nouns chosen randomly from the Paivio, Yuille, and Madigan (1968) norms. Of these 120 nouns, 60 were randomly chosen to be on the list to be presented to all of the *Ss*; the other 60 nouns served as distractors for the *Ss* in the recognition conditions.

Each of the 60 nouns on the list to be presented was mounted as a slide. The slides were then randomly arranged to determine order of presentation for all of the *Ss*. Presentation was via a Kodak Carousel projector (model 850) with an external timer. Answer sheets were prepared for *Ss* in the recognition and recall conditions. For the recall conditions, the answer sheets contained 2 columns of 30 blank lines. At the top of the answer sheet was an instruction asking *S* to write down, in any order he desired, as many words as he could remember from the list he had been presented. For the recognition conditions, the 60 presentation and 60 distractor items were randomly arranged and typed in 4 columns of 30 items each. The instruction at the top of the answer sheet asked *S* to put a line through all of the words he thought were contained in the list he had been presented.

A 2 × 2 factorial design, with both factors varying between groups of *Ss*, was used. The first factor was whether the *S* was given standard learning instructions or instructions to group items. The second factor was whether *S* was given a recognition or recall test. The 4 resulting conditions were labeled standard-recall, standard-recognition, grouping-recall, and grouping-recognition.

The *Ss* were 57 volunteers from introductory psychology classes. They were tested in groups of 1-3 with all *Ss* in a group receiving the same instructional manipulation. There were 14 *Ss* in each condition except the standard-recognition condition, which had 15 *Ss*.

The *Ss* in all conditions were told that they would be presented with 60 nouns that would be familiar to them. Upon seeing each noun, they were to try to remember it as best they could. They were also told that after seeing all 60 nouns, they would be tested on how well they remembered them, although they were not informed of the nature of the test.

The *Ss* in the grouping conditions were given additional instructions about learning the list. These instructions told the *Ss* to group the words in input order by using visual imagery. They were to use 5 or fewer items in each group, and they were given an example of how to group a set of items.

If the *Ss* had no questions, the list was then presented. Rate of presentation was 4 sec. per item.

[1] [illegible] University of British Columbia [illegible]

[2] Now at Rockefeller University.

[3] Requests for reprints should be sent to Michael S. H[illegible], Northwestern University, Evanston, Illinois 60201.

Immediately following presentation and according to their conditions assignment, the *S*s were handed a recognition or a recall answer sheet. The *S*s were told that the instructions were at the top of the sheet and that they would have 4 min. to complete the task.

Following Kintsch (1968), the score for *S*s in all conditions was the difference between the number of correct and the number of incorrect responses. Means for these scores were 45.29, 42.27, 20.71, and 16.29 for the grouping–recognition, standard–recognition, grouping–recall, and standard–recall conditions, respectively. Each *S*'s score was transformed using a square root transformation, and the transformed data were subjected to analysis of variance. The results of the analysis indicated that the recognition scores were higher than the recall scores, $F(1, 53) = 191.60, p < .001$. The grouping instruction resulted in significantly improved recall performance, $F(1, 53) = 4.30, p < .05$, but not significantly improved recognition performance, $F < 1$.

A second analysis was performed on the data of the recall conditions to assure that the superiority of the grouping–recall condition could be attributed to the instructional manipulation. As the *S*s in the grouping conditions were told to form groups of 5 words according to input order, it was predicted that the protocols of the *S*s in the grouping–recall condition would show more clustering according to input order than those of the *S*s in the standard–recall condition. To test this prediction, the presentation list was treated as if it contained 12 categories of 5 words each. Each category was a group of 5 words presented adjacently, starting with Words 1–5, 6–10, etc. Then each *S*'s recall protocol was analyzed according to Frankel and Cole's (1971) adaptation of the runs test, and a z score was determined for each *S*. The mean z scores for the grouping–recall and standard–recall conditions were 3.53 and 1.93, respectively. Both means indicated significantly more clustering according to input order than would be expected by chance, $z = 13.20$ for grouping–recall and $z = 7.21$ for standard–recall, both $ps < .001$. However, the grouping–recall condition showed significantly more clustering according to input order than the standard–recall condition, $z = 4.24, p < .001$. Thus, it can be concluded that the grouping–recall *S*s were organizing in accordance with the instructions.

Experiment II. Rundus (1970) had *S*s rehearse aloud during list presentation and found that items that had been rehearsed together were likely to be recalled together. Thus, if *S* could be prevented from rehearsing items together, it is possible that his level of organization would be depressed. In Experiment II *S*s in the experimental groups were required to rehearse aloud the item being presented, thus reducing the opportunity to rehearse 2 or more items together (see, e.g., Fischler, Rundus, & Atkinson, 1970). The *S*s in the control groups were not required to rehearse aloud and were not asked to use any particular rehearsal strategy.

The stimuli were 480 words, 3 from each of 160 conceptual categories in the Battig and Montague (1969) and Shapiro and Palermo (1969) norms. The words were high associates to the category names. One word from each category was chosen to be on the lists to be presented; the remaining 2 words from each category served as distractors for the recognition conditions.

The 160 words to be presented were typed on 3×5 in. cards. The cards were then randomly divided into 4 lists of 40 words each. Within each list, the cards were randomly arranged to form a presentation order for that list.

Recall answer sheets were prepared in the same manner as in Experiment I, except there were only 20 blank lines in each of 2 columns. Four separate recognition answer sheets were prepared, one for each of the 4 lists. A recognition answer sheet for a given list contained a random arrangement of the 40 words on that list and the 2 distractor words from the categories of those 40 words. The recognition answer sheets were arranged in 4 columns of 30 words each, and the instructions at the top of the recognition answer sheets were the same as those used in Experiment I.

Like Experiment I, Experiment II could be conceptualized as a 2×2 factorial design. The first factor was rehearsal conditions, silent or overt, and the second factor was type of test, recognition or recall. Thus, there were 4 conditions which were labeled silent–recognition, overt–recognition, silent–recall, and overt–recall.

However, unlike Experiment I, the 4 conditions were varied within *S*s so that each *S* served in each of the 4 conditions. Each *S* was presented each of the 4 lists and served in one of the 4 conditions on that list. To counterbalance order of conditions and presentations of lists, a Greco-Latin square design was used. The Greco-Latin square resulted in 4 arrangements of lists and conditions for presentation.

The *S*s were 20 volunteers from introductory psychology classes. Five *S*s served in each of the 4 Greco-Latin square arrangements. The *S*s were tested individually.

Each *S* was read 2 sets of instructions, one set before he served in each of the 2 rehearsal conditions. In neither set of instructions was *S* informed of the type of test which would follow the presentation of the list.

For both sets of instructions, *S* was told that he would be presented 40 words that were familiar to him. He was shown a deck of cards on which one of the lists of words was typed and was told that he would be presented the list, one word at a time, by *E* turning over the cards. The *S* was told that a card would be turned over every 5 sec. and that he should try to remember as many words as he could. In addition, for the silent rehearsal conditions, *S* was told he could study the words in any way that he wished. For the overt rehearsal conditions, *S* was told to repeat aloud the word he was currently being presented. He was to repeat the word throughout the 5 sec. that he saw it and repeat it at a 1-sec. rate.

After S was read the set of instructions appropriate for his first 2 lists, the first list was presented. The click of an electric timer paced E's presentation. Immediately following the completion of first-list presentation, S was handed an answer sheet appropriate to the condition and list under which he was being tested. He was given 2 min. to complete the answer sheet. Then he was told that he would be presented another list and he was to rehearse this list in the same way as he rehearsed the first list. Following the test of the second list, S was read instructions appropriate for rehearsal of the third and fourth lists. The third and fourth lists were then presented and tested.

The *Ss*' protocols were scored as in Experiment I. Means for these scores were 27.60, 28.60, 12.75, and 10.00 for the silent–recognition, overt–recognition, silent–recall, and overt–recall conditions, respectively. Again the data were transformed using a square root transformation, and analysis of variance was used to examine the transformed data. The results of this analysis indicated that recognition scores were higher than recall scores, $F\ (1, 48) = 905.04$, $p < .001$. The instruction to rehearse overtly resulted in a significant decrement in recall performance, $F\ (1, 48) = 18.57$, $p < .001$, but no significant decrement in recognition performance, $F < 1$. Thus, Experiment II has replicated Experiment I by showing an effect of organization in recall but not in recognition.

Discussion. The data from the 2 experiments clearly suggest that the instructional manipulations affected recall but not recognition performance. As the present experiments manipulated organization without changing list members or order of presentation, the results are an alternative form of evidence that organization does not affect recognition accuracy.

The present data are also supportive of a 2-process theory of recognition and recall, such as that offered by Kintsch (1970). A 2-process theory includes the argument that there are qualitative, rather than quantitative, differences between recognition and recall. One-process theories, such as the assertion that the superiority of recognition as compared to recall performance lies in the greater "strength" necessary for recall of those items, predict that experimental manipulations would have the same effect on recognition and recall. The present data, which showed that instructional manipulations of organization affected recall but not recognition performances, are an example of the type of data which Kintsch (1970) cites as support for his 2-process theory.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt. 2).

Bower, G. H., Lesgold, A. M., & Tieman, D. Grouping operations in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 481–493.

Bruce, D., & Fagan, R. L. More on recognition and free recall of organized lists. *Journal of Experimental Psychology*, 1970, **85**, 153–154.

Fischler, I., Rundus, D., & Atkinson, R. C. Effects of overt rehearsal procedures on free recall. *Psychonomic Science*, 1970, **19**, 249–250.

Frankel, F., & Cole, M. Measures of category clustering in free recall. *Psychological Bulletin*, 1971, **76**, 39–44.

Jacoby, L. J. Effects of organization on recognition memory. *Journal of Experimental Psychology*, 1972, **92**, 325–331.

Kintsch, W. Recognition and free recall of organized lists. *Journal of Experimental Psychology*, 1968, **78**, 481–487.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Pt. 2).

Rundus, D. An analysis of rehearsal processes in free recall. (IMSSS Tech. Rep. 159) Stanford, Calif.: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1970.

Shapiro, S. I., & Palermo, D. S. Conceptual organization and class membership: Normative data for representatives of 100 categories. *Psychonomic Monograph Supplements*, 1969, **3**(Whole No. 43).

(Received November 10, 1972)

Journal of Experimental Psychology
1[illegible]4, Vol. 102, No. 3, 520 522

CONCRETENESS OF PEG WORDS IN TWO MNEMONIC SYSTEMS

DENNIS J. DELPRATO[1] AND ELIZABETH J. BAKER

Eastern Michigan University

Two experiments provide evidence that (*a*) concreteness of peg words facilitates recall in the one–bun mnemonic system as well as in a paired-associate (PA) system, and (*b*) the one–bun system, in which peg-word retrieval cues are not explicitly presented at study and recall, and the PA system, in which peg words are explicitly presented, yield comparable results with respect to the effect of peg-word concreteness on recall. These findings are consistent with current views of imagery.

The one–bun (Bugelski, Kidd, & Segmen, 1968) and paired-associate (PA) techniques (Wood, 1967) have been used to study the role of imagery in verbal learning and memory. In the former approach, *S* first memorizes a list of rhyming numbers and peg words (e.g., ONE–BUN, TWO–SHOE). Once the rhyme list has been learned, *S* is instructed to learn new words in ordinal position of presentation by associating them with images evoked by the corresponding rhyming peg words. The PA mnemonic system involves presenting *S* with the list of peg words during study and recall trials; as with the one–bun system, *S* is instructed to integrate to-be-learned words with images aroused by the pegs. Both techniques are basically typical PA tasks on which response terms (to-be-learned words) are learned by associating them with stimulus terms (pegs). The approaches differ because the peg words are explicitly supplied by *E* in the PA technique, whereas they must be memorized by *S* in the one–bun technique.

According to contemporary views of imagery and the conceptual-peg hypothesis (Paivio, 1969), the higher the concreteness of stimulus items, the [illegible] function as mediators of associative learning. Given this basic theory, the empirical evidence of the capacity of concrete nouns to readily evoke images, and the documented facilitative effect of concreteness with the PA system (Paivio, 1969), concreteness of pegs in the one–bun mnemonic [illegible]

This latter finding led Paivio (1968, 1969) to [illegible] that the effect of peg concreteness in the one–bun system is attenuated because the pegs are [illegible] it; with instructions to use imagery, abstract [illegible] imagery as concrete pegs. The situation with the one–bun system is presumed to differ from that with the PA system, which consistently shows concreteness effects, due to the implicit nature of the pegs (retrieval cues) in the former and their explicit nature in the latter system. In other words, Paivio assumes a fundamental difference in the functional effectiveness of identical pegs that is based upon their nominal presence or absence, even though *S* can demonstrate they are in memory by his recall of the pegs. This assumption is not supported by experiments showing that well-learned retrieval cues can be just as effective in mediating recall when physically absent as when physically present (e.g., Delprato & Garskof, 1969); these findings are not surprising if activation of the memory trace is critically dependent upon *S*'s functional encoding of the nominal stimulus, since mere external absence of the stimulus should not influence its encoding when it is in memory.

On the basis of considerations such as the above, we submit that it would be more consistent with current imagery and encoding theory if peg concreteness were a significant factor in the one–bun technique than if it were not. Therefore, Experiment I was designed as a partial replication of Paivio's (1968) study.

EXPERIMENT I

Method. The *Ss* were 84 male and female undergraduates.

The concrete and abstract rhymes were identical with those used by Paivio (1968), i.e., ONE–BUN, TWO–SHOE, etc., and ONE–FUN, TWO–TRUE, etc., respectively.

Two 10-item word lists (A and B) to be recalled were composed of nouns from the Paivio, Yuille, and Madigan (1968) norms. The mean imagery ratings, associative meaningfulness, and Thorndike-Lorge frequencies for Lists A (first value of each pair) and B (second value) were 6.31, 6.30; 5.98, 5.98; 27.8, 22.7, respectively.

The *Ss* were individually tested and were randomly assigned to 1 of 3 conditions (one–bun, one–fun, and control), with 28 *Ss* in each. All *Ss* were given a single study and recall trial on each [illegible]. Half of the *Ss* in each group received List A first, the other half received List B first. The instructions for the first list did not include mnemonic instructions and were identical for all *Ss*. The *Ss* were read the list of 10 words at a 4-sec. rate, with the words preceded by the numerals 1–10. After the list was read once, the numerals alone were read in a prepared random

[1] Requests for reprints should be sent to Dennis J. Delprato, Department of Psychology, Eastern Michigan University, Ypsilanti, Michigan 48197.

order and S was given 10 sec. per numeral to recall the corresponding item. One point was given for each word recalled to the appropriate number.

After S recalled the first list, those who were in the rhyme conditions were told that they were going to go through the same thing again with another list. This time, however, *Ss* were told that an easy technique for learning the 10 words required that they first learn a simple rhyme, then use mental images to picture the object represented by the word in the list in relation to the word that rhymed with the particular number given with the word. The *Ss* were then read the rhyme (abstract or concrete) and tested on it until they could recite it 5 successive times without error. Examples appropriate to the rhyme condition were given. The *Ss* in the control condition did not receive mnemonic instructions between lists. Instead, E engaged S in casual conversation for approximately the same length of time required to give the mnemonic instructions to *Ss* in the experimental conditions.

Results and discussion. The mean recall scores on the first list were 3.89, 3.96, and 4.00 for the control, one-fun, and one-bun conditions, respectively. The means on the second list were 4.96, 5.29, and 6.36 for control, one-fun, and one-bun, respectively. The data were analyzed by a 3×2 (Groups $\times$ Lists) factorial analysis of variance with lists as a within-*Ss* variable. The between-groups effect over both lists was not significant ($F = 1.32$) but the list effect, $F\ (1, 81) = 39.85$, $p < .001$, and the Groups $\times$ List interaction, $F\ (2, 81) = 3.25$, $p < .05$, did reach significance. The significant interaction indicates differential improvement over the 2 lists in the different groups; analysis of the simple main effect of the list variable showed that all 3 groups did improve significantly from the first to the second list, $Fs\ (1, 81) > 4.92$, $p < .05$.

Individual comparisons via Duncan's range test among the group means on each list yielded results consistent with the prediction that peg concreteness is a critical variable in the one-bun technique. The groups were statistically equivalent prior to differential treatment ($p > .05$). On the critical second list, the abstract-peg one-fun rhyme was not more facilitative than the control treatment ($p > .05$). The one-bun group was superior to the control group ($p < .05$) and, in contrast with the results of Paivio (1968), the concrete-peg one-bun rhyme led to significantly higher recall scores than the abstract-peg one-fun rhyme ($p < .05$).

Paivio's (1968) failure to obtain an effect of peg [illegible] questioned by Experiment I), coupled with repeated demonstrations of the effect of this variable with PA learning, suggested to him that the 2 systems produce different results with respect to peg-word concreteness, i.e., the Mnemonic System $\times$ Peg Concreteness interaction. Wood and Bolt (1970) tested this implication with the same materials in the 2 systems and did not find the predicted interaction, i.e., comparable results were found with the 2 systems. However, since they did not find an effect of peg concreteness with either the one-bun or the PA system with their materials, Wood and Bolt's finding of equivalent effects of peg concreteness in the 2 systems is inconclusive in that they did not show the expected effect of concreteness with the PA technique.

Comparable recall differences between lists learned with concrete and abstract pegs in one-bun and PA systems are predicted if functional encoding of retrieval cues is not differentially affected by the mere presence or absence of the nominal stimulus at the time of study and test. While the results of Experiment I indirectly support this prediction, Experiment II was designed to test it directly. Further, the role of peg concreteness in the one-bun system, theoretically important in itself, was again attacked.

Experiment II

Method. The *Ss* were 64 undergraduates of both sexes. Two mnemonic systems (one-bun vs. PA) were factorially combined with 2 levels of peg concreteness (concrete vs. abstract) to form 4 independent conditions ($n = 16$/group).

The materials were identical to those of Experiment I. In the present study, the abstract and concrete peg words became the abstract and concrete stimulus words for the PA conditions, therefore, both stimulus and peg words are referred to as pegs.

All *Ss* were given 2 study and test trials on a single list, with mnemonic instructions prior to the first trial. Half of the *Ss* in each condition received List A and half List B. The study-test procedure on both trials for *Ss* in the one-bun and one-fun conditions was the same as that used with the second list in Experiment I.

The procedure for PA *Ss* was as similar as possible to that of one-bun *Ss*. The PA *Ss* were not required to learn the rhyme, but they were given instructions corresponding to those given to one-bun *Ss* on the use of imagery for study and recall. The E read the peg words to PA *Ss* on study and recall trials in place of the numbers read to one-bun *Ss*. On recall trials, E read PA *Ss* the peg words in a prepared random order (a different order on each trial) that corresponded to the random order in which the number cues were read to one-bun *Ss*.

TABLE 1

Mean Recall Scores, Experiment II

Trial	One bun fun		Paired associate	
	Abstract peg	Concrete peg	Abstract peg	Concrete peg
1	4.50	5.69	4.75	6.31
2	8.06	8.50	6.94	8.75

Results and discussion. The mean number of words correctly recalled to the pegs on each trial by the 4 groups is presented in Table 1. Groups receiving concrete pegs in both mnemonic systems had higher recall scores than groups with abstract pegs. The only significant effects in a Mnemonic System × Peg Concreteness × Trials analysis of variance were those of trials, F (1, 60) = 143.13, $p < .001$, and concreteness, F (1, 60) = 6.11, $p < .025$. The nonsignificant Mnemonic System × Peg Concreteness interaction, F (1, 60) = 1.02, supports our suggestion that the functional encoding of stimulus pegs of varying concreteness would not be differentially affected by their mere nominal presence or absence at the time of study and test.

In conclusion, it appears that peg concreteness is a critical factor in both one–bun and PA mnemonic systems. These findings are consistent with contemporary imagery theory and the conceptual-peg hypothesis (Paivio, 1969).

REFERENCES

Bugelski, B. R., Kidd, E., & Segmen, J. Image as a mediator in one-trial paired-associate learning. *Journal of Experimental Psychology*, 1968, **76**, 69–73.

Delprato, D. J., & Garskof, B. E. Free and modified free recall measures of response recall and unlearning. *Journal of Experimental Psychology*, 1969, **81**, 408–410.

Paivio, A. Effects of imagery instructions and concreteness of memory pegs in a mnemonic system. *Proceedings of the 76th Annual Convention of the American Psychological Association*, 1968, **3**, 77–78. (Summary)

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–262.

Paivio, A., Yuille, J. C., & Madigan, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968 **76**(1, Pt. 2).

Wood, G. Mnemonic systems in recall. *Journal of Educational Psychology Monograph*, 1967, **58**(6, Pt. 2).

Wood, G., & Bolt, M. Type of instruction, abstractness, and mnemonic system. *Psychonomic Science*, 1970, **21**, 91–92.

(Received March 24, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 522–525

PROBABILITY OF CONDITIONED RESPONSES AS A FUNCTION OF VARIABLE INTERTRIAL INTERVALS[1]

KARL HABERLANDT,[2] KEVIN C. HAILS, AND ROBERT LEGHORN

Trinity College, Hartford

Rabbits were trained and extinguished in classical eyelid conditioning with variable intertrial intervals. Responding in each rabbit was recorded as a result of the interval immediately preceding a trial. Conditioned responses were more likely in early acquisition and less likely in early extinction when trial $n + 1$ followed trial n by a shorter rather than a longer interval. These results confirm predictions of stimulus fluctuation and 2-store theories of memory which suggest that events experienced on a trial, whether in acquisition or extinction, are more available and generally influence performance more the sooner a subsequent trial is given.

The present experiment examined the probability of conditioned responses (CRs) on trial $n + 1$ as a result of the interval since trial n in light of predictions based on Atkinson and Shiffrin's (1968) memory theory and Estes' (1955a,b) stimulus fluctuation theory. Both theories expect that a CR should be more likely in early acquisition and less likely in early extinction when trial $n + 1$ is preceded by a relatively short rather than a long intertrial interval (ITI). Atkinson and Shiffrin suggested that a CR occurs on trial $n + 1$ when information about the preceding trial is in the short-term store (STS). If the information is not in STS, a search will be made for it in the long-term store (LTS). Early in training, such information is found only in STS, where it decays relatively quickly. Accordingly, whether in acquisition or extinction, S should recall events from trial n better at shorter than at longer ITIs. Similarly, Estes based his prediction on the assumption that stimulus elements of trial n are better available on trial $n + 1$ when the ITI is relatively short, thus allowing little opportunity for stimulus fluctuation.

Estes (1955, pp. 375–376) cautioned that stimulus fluctuation factors may be masked by more transient posttrial processes. Previous work on ITI effects confirmed Estes' caution and showed that brief posttrial processes may depress or enhance subsequent responding temporarily. Prokasy (1965) and Prokasy and Papsdorf (1965) investigated the probability of CRs as a function of the preceding ITI within the context of evaluating opposing predictions from Estes' (1955) theory and Hull's (1951) reactive inhibition concept. The experiments studied the acquisition of eyelid responses in

---

[1] This report is based on a paper read at the meeting of the American Psychological Association, Honolulu, September 1972.

[2] Requests for reprints should be sent to Karl Haberlandt, Department of Psychology, Trinty College, Hartford, Connecticut 06106.

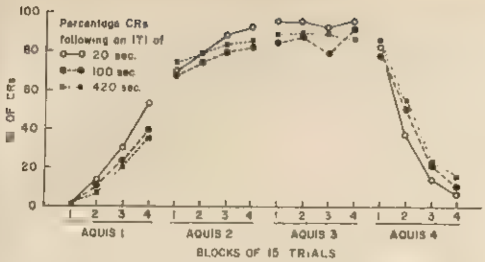


FIGURE 1. Mean percentage of conditioned responses (CRs) following each of 3 intertrial intervals (ITIs) in 3 acquisition sessions and 1 extinction session.

a classical conditioning situation using variable ITIs within *Ss*. Prokasy used ITIs of 2, 4, 6, 8, 10, 12 and 14 sec. in human eyelid conditioning, and Prokasy and Papsdorf used ITIs of 15, 20, 25, 30, 40, and 45 sec. in rabbit eyelid conditioning. Both studies reported that CRs tend to increase with longer ITIs, reflecting the dissipation of reactive inhibition, as Hull had proposed. Inspection of Prokasy's and Prokasy and Papsdorf's data suggests that the probability of CRs increases most between the 2 shortest ITIs, remaining essentially stationary thereafter. Apparently, the response decrement is temporary, and the influence of reactive inhibition decreases or is balanced by stimulus fluctuation after about 20 sec. following trial n. Temporary response enhancement (up to 5 sec.) was reported by Prokasy in those conditions where only 2 ITIs, a short and a long one, were randomly distributed. According to Prokasy, "*S* learns to pay attention immediately following a trial and then in the absence of another trial, relaxes and is subsequently not ready for one when it does occur [p. 471]." Temporary response enhancement, lasting up to 10 sec., was observed by Landauer (1969) in a free-responding situation and was taken as evidence for "posttrial hyperexcitability" which may be initiated by the unconditioned stimulus (UCS). In addition to the choice of relatively short ITIs, previous research may have made it difficult to detect ITI effects occasioned by stimulus fluctuation in another way. That is, ITIs are assumed to affect responding differentially in early acquisition. Later in training, events from trial n are expected to be in STS and LTS and equally available, regardless of the length of the ITI, and hence CR probability should be invariant (Prokasy, p. 471). The studies mentioned above, however, do not detail percentage of CRs during early acquisition.

The present study sought to evaluate the expectations of Estes' (1955) and Atkinson and Shiffrin's (1968) theories by minimizing transient posttrial processes through the choice of relatively long ITIs (20, 100, and 420 sec.) and by observing responding from early acquisition through extinction.

Method. Data were collected from 22 male New Zealand white rabbits in acquisition and from 21 of these rabbits in extinction.

A 3,160-Hz. tone of 88 db. as measured at the position of *S*'s head (General Radio sound pressure meter, Scale 20 kHz.) was used as the conditioned stimulus (CS) and an 8-ma. shock delivered through 2 stainless-steel sutures (ga. #32) across the eye was used as the UCS. Eyelid responses were measured by means of a microtorque potentiometer affixed to *S*'s head and an Offner dynograph. During experimentation, each *S* was restrained in a 50.8 × 17.8 × 14.0 cm. Plexiglas box which was located in a 66.0 × 63.5 × 53.5 cm. sound-attenuated air-ventilated chamber. After *Ss* had been habituated for 4 hr. to the restraining box and the experimental chamber, they received 3 acquisition sessions and then 1 extinction session each, separated by 18 hr. On each of 61 acquisition trials per session, the CS was presented for 600 msec., overlapping with the UCS for the last 100 msec. On extinction trials, the CS was presented alone. Trials were separated throughout by an average ITI of 180 sec. The individual ITIs of 20, 100, and 420 sec. each occurred 20 times in each session, their occurrence following a predetermined order with the restrictions that each ITI occur twice within 6 trials and that during each block of 18 trials, the ITIs follow each other with equal frequency. In acquisition, all *Ss* received the same ITI sequence. In extinction, *Ss* received the same sequence as in acquisition, except that the sequence was started at an ITI of 20 sec. for 7 *Ss*, at 100 sec. for 7 other *Ss*, and at 420 sec. for the remaining *Ss*.

Results. The CRs following the 3 different ITIs were analyzed separately within *Ss*. Figure 1 presents the mean percentages of CRs following each of 3 ITIs during 3 sessions of acquisition and the extinction period. In Session 1 of acquisition, *Ss* responded significantly more following an ITI of 20 sec. than following the longer ITIs, Friedman $\chi^2(2) = 9.30$, $p < .01$. Specifically, CRs following an ITI of 20 sec. were more likely than CRs following ITIs of 100 sec., $t\ (21) = 1.93$, $p < .10$, and 420 sec., $t\ (21) = 2.45$, $p < .05$. There was no significant difference in responding following the latter 2 ITIs.

Analyses of runs of consecutive CRs for each *S*, regardless of the ordinal number of trials on which CRs occurred, showed that of the *first* 10 CRs, 40.9% followed an ITI of 20 sec., 31.8% followed

TABLE 1

MEAN PERCENTAGE CONDITIONED RESPONSES ON TRIAL $n + 1$ AS A FUNCTION OF INTERTRIAL INTERVALS AND PRESENCE OR ABSENCE OF A CR ON TRIAL n

Conditioned responses on trial n	Intertrial interval (in sec.)		
	20	100	420
Session 1 of acquisition[a]			
Present	83.8	69.8	59.4
Absent	16.5	7.3	6.7
Extinction[b]			
Present	71.9	82.5	75.7
Absent	7.4	16.5	26.2

[a] $n = 15$.
[b] $n = 21$.

an ITI of 100 sec., and only 27.3% followed an ITI of 420 sec., $\chi^2(2) = 7.66$, $p < .05$, although there was an equal opportunity (33.3%) for CRs to follow each ITI equally often. Later in acquisition, the percentages converged, such that out of the first 40 CRs, 35.1% followed an ITI of 20 sec., 32.5% followed an ITI of 100 sec., and 32.4% followed an ITI of 420 sec. Table 1 summarizes percentage CRs on trial $n + 1$ following each of the 3 ITIs as a function of the presence or absence of a CR on trial n. (Prokasy & Papsdorf, 1965). Percentages are shown only for Session 1 of acquisition and extinction, because most *S*s had missing data in some cells. For the same reason, only 15 *S*s were included for Session 1 of acquisition. As in the Prokasy and Papsdorf experiment, CRs were more likely on trial $n + 1$ when a CR was present on trial n than when there was no CR, F (1, 14) = 84.45, $p < .01$. This finding was a reflection of the tendency of CRs to occur on adjacent trials rather than on separate trials. Similarly, the ITI effect was significant, F (2, 28) = 3.88, $p < .05$. However, unlike Prokasy and Papsdorf's study, CRs were more likely following the short than the longer ITIs. This result obtained regardless of whether or not a CR occurred on trial n as indicated by the absence of a CR Presence vs. Absence × ITI interaction, F (2, 28) = 1.22. In Sessions 2 and 3 of acquisition there appeared to be somewhat more responding following the short than the longer ITIs, but the difference was not significant.

In extinction, the pattern of responding was opposite to that found in early acquisition, i.e., CRs were more likely following the 2 longer ITIs than following the short ITI, as indicated by a significant main effect, F (2, 40) = 5.20, $p < .01$. Specifically, CRs were significantly less likely after an ITI of 20 sec. than after ITIs of 100 sec., t (20) = 2.58, $p < .02$, and 420 sec., t (20) = 2.88, $p < .01$. There was no significant difference in responding following the latter ITIs. Of the first 10 failures to respond, analyzed separately for each *S*, 40.5% followed an ITI of 20 sec., 31.9% followed an ITI of 100 sec., and only 27.6% followed an ITI of 420 sec., Friedman $\chi^2(2) = 3.90$, $p < .15$. Analysis of variance of percentage CRs on trial $n + 1$ as a function of ITI and presence vs. absence of a CR on trial n revealed a significant CR presence vs. absence effect, F (1, 20) = 181.98, $p < .01$, and a significant ITI effect, F (2, 40) = 8.86, $p < .01$. The significant CR Presence vs. Absence × ITI interaction, F (2, 40) = 4.12, $p < .05$, shows that ITIs affected the occurrence of CRs on trial $n + 1$ differentially, depending on whether or not a CR was present on trial n. That is, percentage CRs on trial $n + 1$ increased between ITIs of 20 and 100 sec., regardless of whether a CR was present on trial n, t (20) = 2.87, $p < .01$, or absent, t (20) = 2.47, $p < .05$. Beyond an ITI of 100 sec., CRs on trial $n + 1$ continued to increase when no CR was present on trial n, t (20) = 1.98, $p < .10$, but they appeared to decline, if not significantly, when a CR was present on trial n, t (20) = 1.54, $p < .20$.

Discussion. The present study found that the length of the ITI preceding a trial influenced the likelihood of a CR early in acquisition and extinction. Early in acquisition, a CR was more likely when the preceding ITI was relatively short than when the ITI was relatively long (100 sec. and longer). On the other hand, in extinction, a CR was more likely when the preceding ITI was relatively long than when the ITI was relatively short.

The results support a stimulus fluctuation theory such as Estes' (1955) and a 2-store theory of memory such as Atkinson and Shiffrin's (1968). These theories suggest that trial events are more available to *S* the sooner a subsequent trial is given. Even if it is assumed for a moment that transient posttrial processes, whether hyperexcitability or reactive inhibition, last for intervals as long as used here, neither process could account for the present data in their entirety. The occurrence of posttrial hyperexcitability could explain the present acquisition results, because it might be argued that the shock UCS had increased the likelihood of a blink response for a brief period. However, the fact that new learning was more prominent at the shorter ITI than at the longer ITIs in extinction cannot be accounted for simply by assuming that a shock UCS temporarily excited *S*. Similarly, while a reactive inhibition construct would predict the higher overall responding at long than short ITIs in extinction, it could not readily explain the higher occurrence of CRs following long than short ITIs when *no* CR was present on extinction trial n, or the higher occurrence of CRs following short than long ITIs in acquisition.

The advantage of Estes' (1955) stimulus fluctuation theory and Atkinson and Shiffrin's (1968) memory theory is their ability to handle the present data with comparable ease and to integrate them with results obtained in human verbal learning. For example, the results reported here invite comparison with results from some continuous paired-associate experiments, especially Bjork's (1966) study. In Bjork's research, *S*s were tested for the recall of paired-associate responses after varying interpresentation intervals (IPIs). Early in training, i.e., before the trial of the last error, the proportion of correct responses was higher the shorter the IPI. Bjork suggests that these short-term effects may be due to the better retention of items in STS with shorter than with longer IPIs.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation, Vol. 2 Advances in research and theory.* New York: Academic Press, 1968.

Bjork, R. A. *Learning and short-term retention of paired-associates in relation to specific sequences of interpresentation intervals.* (Tech. Rep. 106) Stanford: Institute of Mathematical Studies in the Social Sciences, Stanford University, 1966.

Estes, W. K. Statistical theory of distributional phenomena in learning. *Psychological Review*, 1955, **62**, 369–377.

Hull, C. L. *Essentials of behavior.* New Haven: Yale University Press, 1951.

Landauer, T. K. Reinforcement as consolidation. *Psychological Review*, 1969, **76**, 82–96.

Prokasy, W. F. Stimulus fluctuation, reactive inhibition, and

time between trials in classical eyelid conditioning. *Journal of Experimental Psychology*, 1965, **70**, 464-472.
PROKASY, W. F., & PAPSDORF, J. D. Effects of increasing the interstimulus interval during classical conditioning of the albino rabbit. *Journal of Comparative and Physiological Psychology*, 1965, **60**, 249-252.

(Received April 2, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 525-528

TEMPORAL SEPARATION IN VERBAL DISCRIMINATION TRANSFER [1]

STANLEY J. PASKO AND EUGENE B. ZECHMEISTER [2]
Loyola University, Chicago

In verbal discrimination transfer: when situational frequency is initially superior for incorrect (I) than correct (C) items, Hintzman and Block have predicted that the degree to which transfer is inhibited will depend on the temporal separation of the original and transfer tasks. As a test of this hypothesis, *Ss* learned a reversal list task wherein designation of C and I items learned in List 1 was reversed in List 2 (*A*–B, A–*B*, where italics indicate C item). Temporal separation between lists (0 or 7 min.) and number of List 1 trials (4 or 8) were factors. Support for the hypothesis was found only for conditions of low List 1 learning. It is suggested that *Ss* obtain information concerning List 1 context and intrapair associations as number of List 1 trials increases. This information may aid transfer discrimination as frequency accrues to List 2 C items.

A serious shortcoming of the frequency theory of verbal discrimination (VD) learning (Ekstrand, Wallace, & Underwood, 1966) has been the failure to predict, completely, transfer performance in those situations wherein situational frequency is initially greater for incorrect (I) than correct (C) VD items. A study by Raskin, Boice, Rubel, and Clark (1968) is of this general type. These investigators examined VD transfer when pairs from List 1 were retained in List 2, with the designation of C and I items reversed in the second list (*A*–B, A–*B*, where italics indicate C item). Positive transfer was observed on the initial List 2 trial; however, performance first deteriorated slightly and then improved on the next few trials, becoming inferior to that of a control group on remaining trials. The frequency theory assumes that on the initial transfer trial situational frequency is greater for List 2 I than C items due to their prior designation as correct in List 1 and that *Ss* can discriminate correctly by choosing the lower level frequency items as correct. However, as learning demands the continued frequency buildup to C items (e.g., saying aloud the C items on the test trials), the initial advantage to I items is lost and a discrimination based on a relative frequency difference between pair members is made difficult and transfer is inhibited. A strict interpretation of the theory suggests that discrimination should fall to chance levels of responding when the point of frequency equality is reached. Although some investigators, such as Raskin et al., have reported a slight deterioration during early List 2 learning, a complete breakdown in discrimination has not been observed. This is evident even when individual pair frequencies have been adjusted for the effect of wrong responding or, to ensure the hypothetical equality point for all pairs simultaneously, when individual pair frequencies have been rigidly controlled through pronunciation instructions (Underwood & Freund, 1970). It has been concluded that, at some point during second list learning, *Ss* make use of attributes other than frequency to aid discrimination (cf. Underwood & Freund, 1970).

Hintzman and Block (1971) have recently suggested that frequency theory need not predict deterioration to chance levels of responding in the relevant transfer cases if *S*'s ability to discriminate recently from remotely accrued frequencies is taken into consideration. These authors reason that if early in acquisition of the transfer list *Ss* ignore recent frequencies (List 2), and as frequency accrues to C items, later ignore older frequencies (List 1), then deterioration to chance would not be expected. Since the ability to separate List 1 from List 2 frequencies is presumably based on the discriminability of "time tags" present on the list items, it can be expected that "the effectiveness of a manipulation, particularly one designed to inhibit verbal discrimination performance, should depend on the temporal separation of the initial and transfer task [p. 305]."

[1] This experiment is part of research submitted by the first author in partial fulfillment of the requirements for the MA degree at Loyola University, Chicago.
[2] Requests for reprints should be sent to Eugene B. Zechmeister, Department of Psychology, Loyola University, 6525 North Sheridan Road, Chicago, Illinois 60626.

The present experiment is a test of the Hintzman and Block (1971) hypothesis as it applies to VD transfer. Specifically, transfer in a VD task was examined after designation of C and I items learned in List 1 was reversed in List 2 (*A*–B, A–*B*) and when List 2 learning followed List 1 learning either immediately (0 min.) or after a short interval (7 min.). A reversal task was chosen since both members of the transfer list were part of List 1 learning and, therefore, differed only in terms of situational frequency. Also, transfer was examined following 2 different degrees of List 1 learning (4 or 8 trials).

Method. Sixty-four 2-syllable nouns varying 20–50 in Thorndike and Lorge (1944) frequency were selected with an attempt at minimizing semantic and orthographic overlap. Under the restriction that pair members not begin with the same letter, 32 words were randomly selected to form 16 pairs for presentation to the experimental groups. The remaining 32 nouns were randomly paired under the above restriction to serve as List 1 for the control group.

In both experimental and control lists, one item of each pair was randomly designated the C term. For the reversal conditions, 2 forms of List 1 were prepared so that each item might be both a C and I item. In List 2 the C and I relationships were reversed in both forms. For the control groups, one form of List 1 was used for all *Ss*. Control *Ss* were then equally assigned to the 2 forms of the reversal list as List 2. Four orders of all the above forms were created and were presented to *Ss* equally often in counterbalanced form.

After all *Ss* received standard VD instructions, the experimental groups received 4 or 8 trials on List 1. The *Ss* were not informed as to the number of List 1 trials or as to the learning of a second list. The C and I items of a single pair were printed in juxtaposition and exposed on a memory drum. Each pair first appeared for a 2-sec. anticipation interval, after which the C term was shown alone for 2 sec. Across the 4 orders, the C term of each pair appeared equally often, but nonsystematically, in the left and right positions. The intertrial interval for each list was 4 sec. The first trial was always a study trial (*S* did not guess). On the second and succeeding trials, *Ss* responded to each pair, having been told to guess if not sure. The List 1 trial numbers given earlier include the study trial. The control group was arbitrarily given 4 trials on List 1 in order to provide warm-up and/or learning-to-learn experience.

After the requisite number of List 1 trials, experimental *Ss* in the 7-min. temporal separation conditions performed a mathematical filler task, while *Ss* in the 0-min. conditions were immediately informed as to the nature of the second list and that the first trial would again be a study trial. The same instructions were repeated to *Ss* in the delay conditions following the math activity. Control *Ss* were informed that the second list was unrelated to the first and, for these *Ss*, List 2 learning followed immediately after List 1. List 2 learning consisted of 11 trials for all *Ss*, with the first trial a study trial.

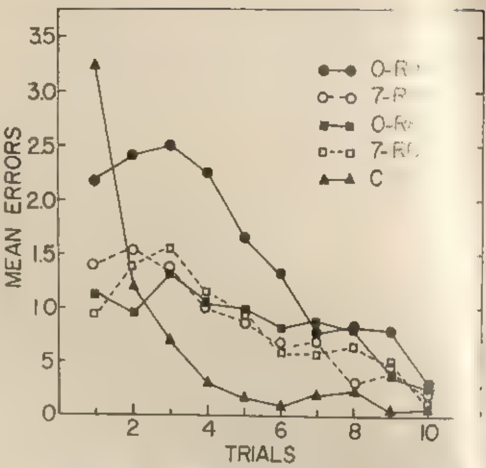


Figure 1. Mean error performance during transfer and acquisition for the experimental and control conditions. (Abbreviations: 0-R4 = 0 min/4-trial reversal list, 7-R4 = 7-min/4-trial reversal list, 0-R8 = 0-min/8-trial reversal list, 7-R8 = 7-min/8-trial reversal list, C = control.)

A total of 120 Loyola University undergraduates participated in the experiment in partial fulfillment of their introductory psychology course requirements. Each of the 5 transfer conditions contained 24 *Ss*. Conditions were randomized in blocks of 5 and assignment of *Ss* was made on the basis of appearance in the laboratory.

Results. Analysis of variance of the total number of errors made on the first 3 List 1 trials indicated that the 4 reversal list conditions did not differ on List 1 learning ($F < 1$). For each experimental condition and for the control condition the mean total errors were as follows: 0-min/4-trial reversal list (0/R4), 8.12; 0-min/8-trial reversal list (0/R8), 8.04; 7-min/4-trial reversal list (7/R4), 6.62; 7-min/8-trial reversal list (7/R8), 8.08; and control, 9.29. Within the experimental conditions, however, the 2 forms of the VD list could not be judged of equal difficulty, F (1, 88) = 4.95, $p < .01$. Nevertheless, the differences between forms of the list did not vary with experimental conditions, F (3, 88) = 2.49, $p > .05$.

Figure 1 shows mean errors made on each of the 10 transfer trials by those *Ss* learning a reversal list and for control *Ss*. Statistical analysis of the reversal list performance indicated reliability for the pattern of results seen in Figure 1. While neither temporal separation nor number of List 1 trials was significant as main effects, the Number of List 1 Trials × Reversal List Trials interaction was significant, F (9, 92) = 3.06, $p < .01$, as was the second-order Temporal Separation × Number of List 1 Trials × Reversal List Trials interaction, F (18, 828) = 2.38, $p <$ 01. These results reflect the initially inferior performance, and subsequent faster rate of improvement, of the 0/R4 condition

relative to the remaining reversal list conditions. The reversal list trials effect was significant, F (3, 92) = 29.33, $p < .01$, but no other effects were reliable. Separate analysis of transfer performance for the 0/R4 and 7/R4 conditions indicated a significant effect for temporal separation, F (1, 6) = 5.19, $p < .05$, as well as for the Tem-...aration × Reversal List Trials interaction, ...1) = 1.94, $p < .05$.

...man–Keuls range test of mean errors on ... transfer trial indicated that the experi-...roups were all significantly different from ...rol group ($p < .05$). Within the experi-...groups, the 0/R4 and 7/R4 groups did not ...nificantly, but the 0/R4 group was sig-... different from the 0/R8 and 7/R8 groups ...). In terms of overall errors, *Ss* in Condi-...R4 were significantly different from the ...aining 4 transfer groups ($p < .05$).

Discussion. The frequency theory of VD learning requires that several predictions be met regarding reversal list performance as a function of number of List 1 trials. For the present task, an increase in number of List 1 trials presumes a larger difference in situational frequency between C and I items at the point of the first transfer trial. If frequency alone serves as a discriminative ...e larger the initial frequency difference be-...nd I pair members, the easier should be discrimination and the greater should be positive transfer on the first reversal list trial. An examination of Figure 1 shows that this expectation is satisfied only when List 2 immediately followed the learning of List 1 (0/R4 vs. 0/R8). The same ...on between conditions differing on number ...l trials, but when transfer list acquisition was delayed (7/R4 vs. 7/R8), shows no overall difference as a function of increasing trials on the original list. A further prediction made on the basis of frequency theory concerns the point during second-list learning at which deterioration of transfer performance should be observed. The larger the initial frequency differential between C and I items, the more transfer trials will be required to approximate an equalization of apparent frequency. Looking again at Figure 1 and comparing groups differing only in number of List 1 trials, there is a slight tendency for delayed deterioration between Conditions 0/R4 and 0/R8, but little difference between Conditions 7/R4 and 7/R8. However, differences are slight and, therefore, the results are equivocal for this prediction. Finally, frequency theory cannot presently account for the difference in performance between conditions differing only in temporal separation of original and transfer lists (0/R4 vs. 7/R4). This latter finding, however, is amenable to predictions made from the multiple-trace hypothesis of Hintzman and Block (1971) and provides evidence that attributes other than frequency (e.g., temporal) are involved in transfer of a discrimination.

An explanation for the fact that temporal separation was effective in reducing transfer list inhibition only for a relatively low degree of List 1 learning may be related to one of several consequences of increasing number of List 1 trials. First, it can be noted that in the present experiment degree of List 1 learning is confounded with temporal separation between the initial trials of the original and transfer lists. It might be logically argued that an increase in the number of List 1 trials would lessen the effectiveness of the present temporal manipulation since the more trials on List 1, the greater is the time associated with List 1 practice and, conceivably, the more discriminable are the time tags. On the other hand, if increasing the number of List 1 trials is correlated with other than temporal changes which can mediate list discrimination, then *Ss* may ignore List 1 frequency, on the basis of these cues, as List 2 frequencies become sufficiently disparate to provide a consistent cue for discrimination performance. Additional cues that can mediate discrimination may be expected to obviate the use of a temporal cue if these added attributes are more dominant in discrimination performance.

Two correlates of increasing number of List 1 trials can be expected. First, as number of List 1 trials increases, so too does *S*'s information regarding List 1 context. If frequency can be assumed to be context related, then it is likely that List 1 frequency is more discriminable after a larger number of List 1 trials. For example, this interpretation would explain the failure of King and Levin (1971) to find inhibition in VD transfer after 8 List 1 trials when between-lists changes involved either designation of a new C item (*A*–B, *C*–B), or both the reversal of List 1 C items and the designation of a new C item (*A*–B, *C*–A). In both these paradigms, situational frequency is greater for I than C items on the initial transfer trial and the frequency theory predicts initial positive transfer and subsequent deterioration on remaining transfer trials. King and Levin's results are in contrast to results seen in Figure 1 where reversal list inhibition after 8 List 1 trials was found. If context aids list discrimination, less inhibition would be expected in those VD paradigms which change list composition between initial and transfer lists (e.g., *A*–B, *C*–B) than in a reversal paradigm wherein changes in original and transfer list learning are not correlated with changes in list composition (*A*–B, A–*B*).

A second correlate of increasing number of acquisition trials is an increase in *S*'s information regarding intrapair associations (e.g., Zechmeister & Underwood, 1969). In terms of reversal list performance, this fact suggests that as frequency accrues to the C items of the transfer list, discrimination may be aided by *S*'s information about intrapair relationships, particularly as equalization of frequency between pair members is approximated.

Finally, all reversal conditions of the present experiment showed a slight deterioration on early List 2 trials and subsequent inhibition on later transfer trials. This finding suggests that *Ss* are not entirely successful at ignoring prior frequency during List 2 acquisition, a fact which tends to

confirm the notion that frequency is the dominant attribute governing discrimination performance and that *S*s have difficulty using less preferred cues.

REFERENCES

EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

HINTZMAN, D. L., & BLOCK, R. A. Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 1971, **88**, 297–306.

KING, M. E., & LEVIN, I. P. Tests of frequency theory in verbal discrimination transfer with varying paradigms and List 1 trials. *Psychonomic Science*, 1971, **25**, 226–228.

RASKIN, D. C., BOICE, C., RUBEL, E. W., & CLARK, D. Transfer tests of the frequency theory of verbal discrimination learning. *Journal of Experimental Psychology*, 1968, **76**, 521–529.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

UNDERWOOD, B. J., & [illegible]ND, J. S. Relative frequency judgments and verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **83**, 279–285.

ZECHMEISTER, E. B., & UNDERWOOD, B. J. Acquisition of items and associations in verbal discrimination learning as a function of level of practice. *Journal of Experimental Psychology*, 1969, **81**, 355–359.

(Received April 4, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 528–530

MOTIVATING EFFECT OF CONTINGENT SELF-REWARD[1]

GISELA E. SPEIDEL[2]

Kamehameha Early Education Project, Honolulu, Hawaii

The motivating effects of contingent self-reward on an intellectual though boring task were studied. The *S*s were instructed to work on simple addition problems for as long as they wished. The *S*s in the self-reward condition presented themselves with sections of a television film contingent upon a fixed-ratio schedule that they had selected from 3 alternatives. An *E*-rewarded group, yoked to the self-reward *S*s, received the movie sections automatically. A control group received rest periods instead of the film. The results partially supported the hypotheses that contingent self-reward can have motivating effects. The self-reward group completed significantly more problems than the control group while there were no significant differences in number of problems completed between the self-reward and the *E*-rewarded groups.

The variables that are relevant to the establishment and occurrence of contingent self-reward have been charted extensively under laboratory conditions (for reviews, see Bandura, 1971; Kanfer, 1970). By comparison, there are relatively few studies seeking to demonstrate that self-reward can have a reinforcing or motivating function. Examination of the most frequently cited studies (Kanfer & Duerfeldt, 1967; Marston & Kanfer, 1963) has revealed methodological problems that make equivocal the conclusion that self-administered reward functions to motivate behavior (Speidel, 1972).

The present experiment was designed to furnish further evidence on the effects of contingent self-reward. Specifically, the study tested whether sections of a movie, self-administered contingently upon the completion of a certain number of addition problems, would serve to increase the number of problems completed and to increase the rate of responding.

[1] This research was submitted in partial fulfillment of the requirements for the PhD degree at the University of Hawaii. The author wishes to thank Roland G. Tharp, W. J. M. Evans, Clifford R. O'Donnell, and John G. Carlson for their helpful critical comments.

[2] Requests for reprints should be sent to Gisela E. Speidel, Kamehameha Schools, Kapalama Heights, Honolulu, Hawaii 96817.

Method. The *S*s were 45 undergraduate psychology students. Fifteen *S*s (7 males and 8 females) were assigned to each of the following groups: self-reward (SR), *E*-rewarded (ER), and no-reward control receiving rest periods. The *S*s in the ER and the control groups were yoked to *S*s in the SR group in terms of the schedule on which they received the reward or the rest period, respectively.

The *S*s were tested individually and were presented with a stack of 70 pages containing 25 simple addition problems each. The problems all had the same pattern of 2 rows and 3 columns of numbers as shown in the example $\begin{array}{r}234\\+867\end{array}$. A panel with 2 buttons and a counter were connected to a central relay system in *E*'s room. One of the buttons activated a cumulative recorder and a counter in *E*'s room and also the counter on *S*'s panel (during the ER and SR conditions). The other button activated a buzzer in *E*'s room; *E* then turned on a videotape recorder (for the SR condition only). A television monitor was placed in the *S* room. A videotaped television detective story served as the reward.

All groups received instructions to work on the

addition problems and to press the "problems completed" button every time they had finished a problem. They were given one page of trial problems for practice. (The rate with which *S*s worked on this page was used as a preexperimental rate measure.)

The SR *S*s were then shown how they could keep track of the number of problems they had completed by checking the counter on their panel. They were instructed to treat themselves to some "movie time" whenever they had finished a certain number of problems by pressing the television button on their panel. The film was briefly described to them and they were shown how to tune the monitor. The *S*s were presented with a card that contained 3 alternative schedules on which they could reward themselves. They were asked to choose one of the schedules (but not to tell *E*). If they found the schedule on which they worked too easy or too difficult they could change schedules, but not more than once.

The ER *S*s were given the same information relating to the counter on their panel, the film, and the television-tuning instructions as SR *S*s. However, they were told that the television would come on automatically each time they had completed a certain number of problems. For each *S*, the schedule on which he was to be rewarded (which was the same as that on which the SR *S* to whom he was yoked had rewarded himself) was stated.

In place of the television instructions the control group was informed that occasionally the loudspeaker would come on and announce break time. During this period they were not to work on the problems. Control *S*s were yoked to SR *S*s in terms of the schedule on which they received the rest period. However, rather than giving the period in a single unit at the completion of the schedule, it was given in 2 parts—one at some point in the middle of the set of problems and the other at the completion of the schedule. This was done so as not to convey to *S*s that the rest period was contingent upon performance.

The final instructions for all 3 groups informed *S*s that the experiment was open ended in length and that they could leave the experiment whenever they wished. When they decided to stop they were to come into *E*'s room and fill out the questionnaire. If *S* had not signaled that he wanted to leave after 900 problems, *E* stopped him. The *S*s were given a questionnaire and asked to rate the movie or the break on a 7-point bipolar scale from -3 (extremely boring) through 0 (the neutral point) to $+3$ (extremely interesting or restful).

Results. The effects of the experimental conditions were measured on 3 dependent variables: (*a*) the number of problems completed; (*b*) the average number of seconds needed to complete a problem, i.e., rate; and (*c*) the decrease in number of seconds per problem from the beginning to the end of the experiment, i.e., increase in rate.

The mean number of problems completed by the control, SR, and ER groups was 527.53, 723.33, and 826.67, respectively. The differences among the means were statistically significant, F (2, 42) $= 7.23$, $p < .005$. A comparison of the means by the Newman–Keuls method showed that both the SR and ER groups completed significantly more problems than the control group; the difference between the ER and control groups was significant at the .01 level; the difference between the SR and control groups was significant at the .05 level. Although the ER group completed more problems than the SR group, this difference was not significant.

Both rate and increase in rate of performance during the experiment showed no significant treatment effects. However, the pretreatment rate correlated positively with rate during the experiment, r (43) $= .78$, $p < .001$, and negatively with an increase in rate, r (43) $= -.61$, $p < .001$. Therefore point-biserial correlations were conducted between treatment conditions and each of these 2 dependent variables with pretreatment rate partialed out. These correlation coefficients were in the predicted direction: The ER group worked faster and increased more in rate during the experiment than the control group while the SR group took an intermediate position. However, the correlations were not significantly different from each other.

The mean rest-period rating given by the control group was .33 and did not differ significantly from the neutral point. The mean movie ratings given by the SR and the ER groups were 1.4 and 2.0, respectively. The movie was rated significantly more positively than the rest period, t (44) $= 4.28$, $p < .001$. The more attractive an *S* rated the movie, the more problems he tended to complete, r (29) $= .28$, $p < .06$.

Discussion. The results on the number of problems completed indicated that self-reward can function to motivate behavior in a manner similar to *E*-presented contingencies. Further support for the conclusion that the difference in performance between the experimental groups and the control group was due to the motivating function of the film, was provided by the findings that the movie was rated significantly more positively than the neutral rest period and that the more attractive the movie was rated, the more problems were completed.

The fact that the other 2 dependent measures showed only tendencies in the predicted direction does not provide any evidence against the present hypothesis. Since not even the ER group exhibited any significant treatment effects, it is concluded that these measures did not constitute a sufficiently sensitive test for the effects of contingent reward whether externally or self-administered. These measures seem to reflect mainly stable individual differences in arithmetic skill, as pretreatment performance accounts for most of their variation.

It is therefore concluded that the present results confirm the findings of Bandura and Perloff (1967) who demonstrated that self-administered tokens could maintain a motor task in children. A broader

data base is thus provided for past and future applied studies using self-reward to effect behavior change.

A warning, however, against overenthusiastic application of self-reward in a clinical setting, at this stage, is appropriate. Experiences gained from conducting pilot work and the present experiment suggest that the motivating effect of contingent self-reward is a delicate one that requires special conditions for its demonstration. Some of the factors to which special attention must be given are the (*a*) other self-control behaviors an individual already has in his repertoire; (*b*) appropriate level of aversiveness of the task on which the motivating effects are to be demonstrated; and (*c*) appropriate level of attractiveness of the reward, so as to ensure motivating effects but not to evoke rule violation. Furthermore, most experiments in self-reinforcement have some demand characteristics for compliance with the rules for contingent self-reward (Bass, 1972). A still unanswered question is whether contingent self-reward can occur without such demand characteristics. Even in the present experiment demand characteristics existed (e.g., the presence of *E* in an adjoining room, wanting to be a good *S*, etc.) that may have been responsible for the maintenance of contingent self-reward, and therefore prevented rewarding oneself before the completion of the criterion behavior. The design of the present study greatly reduced the potential operations of such demand characteristics in that (*a*) *S*s were given a choice of alternative work schedules and the opportunity to change schedules; (*b*) they monitored the number of problems they had completed on their counter; and (*c*) they received no signal to indicate that they had reached their criterion for self-reward and they consumed the reward immediately and did not have to exchange tokens for a reward at the completion of the experiment. The last 2 features had been present in the Bandura and Perloff (1967) study.

In light of the above discussion the conclusion permissible from the present study and Bandura and Perloff's (1967) study is that under circumstances that insure compliance with the rule of contingent self-reward, self-administered rewards can be shown to motivate and maintain behavior in a manner similar to rewards controlled by another person.

REFERENCES

BANDURA, A. Vicarious and self-reinforcement processes. In R. Glaser (Ed.), *The nature of reinforcement*. New York: Academic Press, 1971.

BANDURA, A., & PERLOFF, G. Relative efficacy of self-monitored and externally imposed reinforcement systems. *Journal of Personality and Social Psychology*, 1967, **7**, 11–116.

BASS, B. A. Demand characteristics as a determinant of self-reinforcement. Unpublished doctoral dissertation, University of Tennessee, 1971.

KANFER, F. H. Self-regulation: Research, issues, and [illegible]lation. In F. H. Kanfer & J. S. Phillips, *Learning foundations of behavior therapy*. New York: Wiley, 1970.

KANFER, F. H., & DUERFELDT, P. H. Motivational properties of self-reinforcement. *Perceptual and Motor Skills*, 1967, **25**, 237–246.

MARSTON, A. R., & KANFER, F. H. Human reinforcement experimenter and subject controlled. *Journal of Experimental Psychology*, 1963, **66**, 91–94.

SPEIDEL, G. E. Reinforcement effects of contingent self-reward. Unpublished doctoral dissertation, University of Hawaii, 1972.

(Received April 6, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 531-533

USE OF RULE 1 AND RULE 2 IN VERBAL DISCRIMINATION TRAINING[1]

SLATER E. NEWMAN,[2] RALPH E. SUGGS, AND CAROL H. AVERITT

North Carolina State University

During verbal discrimination training, each *S* in Experiment I was told on study trials which item of each pair was *right* or *wrong*. Appropriateness of Rule 1 (select the more frequent item of the pair) and Rule 2 (select the less frequent item of the pair) was manipulated by having *S* designate on test trials the same item of the pair designated by *E* on study trials (Rule 1 thus being the more appropriate) or the other item of the pair (Rule 2 being the more appropriate). In Experiment II, appropriateness of Rules 1 and 2 was manipulated by having *S* designate on the test the same item of the pair *E* had instructed him to learn during study trials or the other item. In both experiments, performance during training was better for *S*s for whom the use of Rule 1 was the more appropriate.

According to the frequency theory of verbal discrimination (VD) learning (Ekstrand, Wallace, & Underwood, 1966), a frequency unit is added to an item each time *S* responds with it. The theory asserts that correct responding during VD training is directly related to the difference in frequency units between items of a pair. According to Ekstrand et al., *S* responds with the more frequent (Rule 1) or less frequent (Rule 2) item of a pair depending on the requirements of the situation.

A transfer experiment by Underwood, Jesse, and Ekstrand (1964) has provided some data on the use of these rules. After each *S* learned a VD list to 3 consecutive perfect trials, he learned a second list. For half of the *S*s, the *right* items from List 1 were still right on List 2 but the *wrong* items were new; for the other *S*s, the *wrong* items from List 1 were still *wrong* on List 2 but the *right* items were new. Performance on the first test of List 2 was almost perfect for the *right S*s (for whom the use of Rule 1 was the more appropriate). The *wrong S*s (for whom the use of Rule 2 was initially the more appropriate) also made few errors on the first test of List 2, though substantially more than *right S*s. (This difference in performance between the *right* and *wrong S*s can be explained by assuming that the frequency difference between old and new items for the *right* group was greater than for the *wrong* group.) These data thus accord with the statement that "either Rule 1 or Rule 2 can be applied depending on the appropriateness of the situation [Ekstrand et al., 1966, p. 570]."

Experiment I

This experiment studied the use of Rule 1 and Rule 2 in a single-list situation. For half of the *S*s, Rule 1 was the more appropriate from the beginning of training; for the remaining *S*s, Rule 2 was the more appropriate. A 2×2 design was used which permitted study of the effects of (*a*) designating an item as *right* or *wrong* on study trials and (*b*) requiring *S* to designate the *right* or *wrong* items on test trials. Thus, on study trials, half of the *S*s were told which item of a pair was *right* and half were told which item was *wrong*. On test trials, half of the *S*s in each treatment were instructed to indicate the *right* item of each pair, and half were instructed to indicate the *wrong* item of each pair. As in the experiment by Ekstrand et al. (1966), *S* was asked to indicate whether the to-be-designated item was first or second rather than to pronounce it, thus presumably eliminating the buildup of frequency units on test trials. The 4 groups were designated RR, RW, WR, and WW (the first letter refers to the item designated on study trials, i.e., R = right, W = wrong; the second letter to the item to be designated by *S* on test trials). All *S*s were given 4 study trials, each followed by a test.

Method. The 20 pairs were those used in the control condition by Ekstrand et al. (1966). For each treatment, 2 lists were constructed so that on study trials each item of a pair was repeated either on List A or on List B (e.g., in Group RW, half of the *S*s heard "queen–fast, queen is right"; the rest heard "queen–fast, fast is right"). During training, 8 different orders of the pairs were used, 4 for study trials and 4 for test trials. Each item appeared first in its pair on half of the study trials and on half of the test trials.

The *S*s were run individually. A tape recorder was used to present the instructions and to present the items on study trials and on test trials. The instructions told *S* which item of a pair would be designated on study trials and which item he was to designate on test trials. On study trials, each item of a pair was presented and *S* was then told which item was *right* (Groups RR and RW) or *wrong* (Groups WR and WW). On test trials, each item of the pair was presented and *S* called out that item of the pair (first or second) that he believed was *right* (Groups RR and WR) or *wrong* (Groups WW and RW). A 4-sec. rate was used

[1] The authors appreciate the help of Harry K. Dorsett in making students at Meredith College available for participation in this experiment, and of Donald H. Mershon who commented on an earlier version of this manuscript.

[2] Requests for reprints should be sent to Slater E. Newman, Department of Psychology, North Carolina State University, Raleigh, North Carolina 27607.

TABLE 1

Mean Correct on Each Trial for Each Treatment: Experiments I and II

Treatment[a]	Trials 1	2	3	4	Total
Experiment I					
RR	14.25	16.55	18.50	19.35	68.65
RW	11.75	14.40	15.50	17.10	58.75
WR	11.85	13.00	15.30	16.45	56.60
WW	13.35	16.35	17.55	18.95	66.20
Experiment II					
RR	12.70	14.95	17.15	17.40	62.20
RW	10.80	13.25	13.90	14.25	52.20
WR	10.75	12.85	14.10	14.45	52.15
WW	12.80	14.50	17.25	16.60	61.15

[a] Abbreviations: R = right, W = wrong; for further explanation, see description in text for each experiment.

on both study and test trials. There were 4 study trials each followed by a test. A 50-sec. interval separated the first study trial from the first test trial. During this interval, *E* reiterated instructions for the test. Subsequent intervals between each study trial and its following test trial were 15 sec. and between each test trial and its subsequent study trial, 4 sec.

There were 80 *S*s, 56 from the introductory psychology course at North Carolina State University and 24 from a class in educational psychology at Meredith College. A balanced Latin square was used in assigning *S*s from each population to the treatment conditions.

Results and discussion. A 2 × 2 × 4 (Study Trial Repetition × Test Response × Trials) analysis of variance was applied to the data for number correct on Tests 1–4. Only the effects of trials, F (3, 228) = 102.34, $p < .001$, and the Study Trial Repetition × Test Response interaction, F (1, 76) = 30.25, $p < .001$, were significant. Application of Duncan's multiple-range test showed that the means for Groups RR and WW each differed ($p < .01$) from the means for Groups RW and WR; but the RR mean did not differ from that for WW ($p > .05$) and the RW mean did not differ from that for WR ($p > .05$). The mean correct for each treatment on each trial is shown in Table 1.

In accord with predictions derivable from the frequency theory, performance during training was affected neither by the item (*right* or *wrong*) designated on study trials nor by the item to be designated on test trials. Performance was affected only by whether the item designated during study trials was the same as (or different from) the item to be designated on the test trials. If it was the same (as in Groups RR and WW, for which use of Rule 1 was the more appropriate), then performance was better than if it was different (as in Groups RW and WR, for which use of Rule 2 was the more appropriate). Examination of the treatment means on each test shows that the between-treatments differences occurred on the first test and remained relatively stable throughout training.

The improvement in performance during training may be taken as evidence that during VD train . . "either Rule 1 or Rule 2 can be applied depending on the appropriateness of the situation [Ekstrand et al., 1966, p. 570]." The present results appear to indicate, however, that where use of Rule 1 is the more appropriate, performance during training (as early as the first trial) may be be[illegible] where the use of Rule 2 is the more appro[illegible]

Experiment II

In explaining the results from Experiment I [illegible] was assumed that the difference in frequency unit[illegible] between items of a pair derived from the f[illegible]quency of their overt occurrence. In Experim[illegible] *right* and *wrong* items of each pair were p[illegible] only once on each study trial, *S* was instru[illegible] prior to each study trial which item (*right* or *wrong*) he was to learn (but not which item he was to designate on test trials), and was then tested for those items or for the other items. If frequency units accrue to an item only as a result of its overt occurrence, all groups should perform at chance level since the number of frequency units for the 2 items of a pair would be equal. Four trials were run to determine whether performance of any of the 4 groups would change with practice.

Method. The method was similar to that used in Experiment I with the following exceptions: (*a*) *S*s were instructed prior to each study [illegible] to learn the *right* (Groups RR and RW[illegible] (Groups WR and WW) item of each pair [illegible] were not told until after the first study trial which item they would have to designate on test t[illegible] (*b*) on study trials, half of the *S*s heard "q[illegible] right, fast is wrong" (List A) and the rest heard "queen is wrong, fast is right" (List B); (*c*) a 3- rather than a 4-sec. rate was used on study and test trials; (*d*) immediately prior to the first test and to each test thereafter, half of the *S*s in each treatment were asked to indicate (by saying "first" or "second") the items they had been instructed to learn (Groups RR and WW) and the remaining *S*s were asked to indicate the other item of each pair (Groups RW and WR); and (*e*) 50 sec. intervened between each subsequent study trial and the test that followed, and 24 sec. between each test and the study trial that followed.

The design was 2 × 2 with 20 *S*s, 12 male and 8 female, in each treatment. All were undergraduates at North Carolina State University. Males and females were separately assigned in counterbalanced order to each of the treatments.

Results and discussion. To determine whether *S*s performed better than chance on Test 1, a t test was done comparing the mean correct (11.76) for all 80 *S*s and the mean expected by chance. The t (78) = 2.79, $p < .01$, may be interpreted as indicating that at the end of the first study trial a discriminable difference in frequency units existed between the items of each pair.

An analysis of variance for number correct on Test 1 showed a significant effect only for the

Study Instructions × Test Response interaction, F (1, 76) = 6.94, $p < .05$. For both main effects, F was smaller than 1.00. Duncan's test showed no significant differences between the 4 treatment means. When, however, the mean for *S*s in Groups RR and WW was compared with the mean for *S*s in Groups RW and WR, the difference was significant at the .01 level.

A separate analysis of variance for number correct on each trial during training showed that only the effects of trials, F (3, 228) = 53.72, $p < .001$, and the Study Instructions × Test Response interaction, F (1, 76) = 23.01, $p < .001$, were significant. The means appear in Table 1. Duncan's test showed that the means for Groups RR and WW differed from the other 2 means but not from one another, and that the RW mean did not differ from the WR mean.

The results from this experiment replicate those from Experiment I in demonstrating that having *S* designate the *wrong* items on the test has the same effect as having him designate the *right* items. These results demonstrate also that instructing *S* to learn the *wrong* items has the same effect as instructing him to learn the *right* items.

The better-than-chance performance on Test 1 may be explained by assuming that *S*s rehearsed covertly those items they were instructed to learn. If they did, then the better performance of *S*s in Groups RR and WW than of *S*s in Groups RW and WR suggests that performance is better where use of Rule 1 is the more appropriate than where use of Rule 2 is the more appropriate. This could occur if, as in these experiments, *S* had little time to respond and if use of Rule 1 involved fewer operations (e.g., "A is more frequent than B, select A") and thus took less time than use of Rule 2 (e.g., "A is more frequent than B, reject A, select B").[3]

An alternative explanation for these results is based on the assumption that *S*s rehearse covertly those items they expect to designate on the test. Since these are the same items that *S*s in Groups RR and WW were instructed to designate on tests, the difference between items of each pair would increase during training, Rule 1 would become easier to use, and this would be reflected in improved performance over trials. For *S*s in the other 2 groups, RW and WR, the results for Tests 2, 3, and 4 are harder to interpret since *S*s were instructed to learn one item of a pair, but were tested for the other. One possibility is that as training proceeded *S*s came to expect that they would have to designate those items they had to designate on previous tests. Thus during Study Trials 2, 3, and 4, *S*s would rehearse those items (rather than those they were instructed to learn) so that Rule 1 would come to be the more appropriate for *S*s in those groups also. The better performance by *S*s in Groups RR and WW would then be due to a greater difference in frequency units between items of each pair in those groups than in the other 2 groups.

The assumption that *S*s rehearse covertly those items they expect to designate on the tests, can also be used to explain the results from Experiment I. Thus Rule 1 could have been the more appropriate for *S*s of all 4 treatments of that experiment, since all *S*s were told at the beginning of training the items they were to designate on the test and there would be more covert rehearsal for those items. The better performance by *S*s in the RR and WW groups would then be due to a greater difference in frequency units between items of each pair in those groups than in the other 2 groups.

[3] Such behavior for Rule 1 *S*s would also result in a difference in the accrual of frequency units for the 2 items of a pair on test trials.

REFERENCES

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, 73, 566–578.

Underwood, B. J., Jesse, F., & Ekstrand, B. R. Knowledge of rights and wrongs in verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 183–186.

(Received April 7, 1973)

Journal of Experimental Psych[illegible]
1974, Vol. 102, No. 3, 534–536

ACQUIRED PLEASANTNESS AS A STIMULUS AND A RESPONSE VARIABLE IN PAIRED-ASSOCIATE LEARNING[1]

ALBERT SILVERSTEIN[2]

University of Rhode Island

Two experiments are reported in which nonsense syllables paired with pleasant pictures (P-paired syllables) were subsequently learned more rapidly than [illegible] not when they served as stimuli—in paired-associate lists. The bulk of the superiority of P-paired syllables was in response availability, and only a smaller effect was found in association learning. Both effects depended upon Ss learning to anticipate responses differing in acquired pleasantness. Superior ratings of pleasantness were given to the P-paired syllables regardless of whether they were learned as stimuli or responses on the list or whether by anticipation training or associative-match training.

Recent demonstrations that nonsense syllables that have been paired with pleasant pictures (P-paired syllables) are subsequently learned faster in paired-associate (PA) lists than are syllables that have been paired with indifferent [illegible] syllables) have been limited to the case in which the syllables are in the response positions of the list (Silverstein, 1966, 1972a, 1972b). The present study [illegible] nonsense syllables placed in stimulus and response positions of PA lists.

The aforementioned studies support the hypothesis that the acquired-pleasantness effect is based largely upon Ss' disposition to implicitly rehearse P-paired syllables more vigorously than I-paired syllables during the response-learning stage. The bulk of the difference between performance with P- and I-paired syllables has been found in the number of trials before the first response (whether it is correct or not), a measure used to index speed of response learning (Underwood & Schulz, 1960). Moreover, instructions to continuously pronounce the P- and I-paired syllables during the learning task removed the superiority of the former (Silverstein, 1972b). Since [illegible] of the stimulus terms are largely [illegible]

syllables in *either* the stimulus *or* response positions of a PA list. The second task of Experiment I consisted of standard anticipation learning, while that of Experiment II consisted of alternating study and test trials of various types.

Experiment I. Thirty-two female and 48 male Ss were first given 6 trials of rating the difficulty of learning 8 pairs of pictures and nonsense syllables. They were told that we wanted to find out how well people can judge the difficulty of learning a pair of items as a "unitary pair" and were instructed to mark 9-point scales ranging from "very easy" to "very difficult" that were printed on each page of a booklet. The members of each pair were shown simultaneously for 10 sec., and there was [illegible] intertrial interval. On each trial, half of [illegible] had the syllable on the left side and half [illegible] syllable on the right side. The syllable for each pair was shown on each side on an equal number [illegible] but S's task was the same regardless of the [illegible] the syllable and picture. In the second task [illegible] and female Ss were each randomly divided [illegible] groups receiving the syllables from Task 1 as either responses to or stimuli for 2-digit numbers for 1[illegible] anticipation trials. Stimulus and response terms were each shown for 5 sec., and there was a 10-sec. intertrial interval. Following this task, all Ss rated the syllables for pleasantness on a 7-point scale.

Materials shown in both tasks were displayed by a Kodak Carousel slide projector at a distance of 6 ft. from S. Different sets of pictures were used for male and female Ss, having previously been rated [illegible] Ss of the appropriate sex. Four pictures in [illegible] list were in the "very pleasant" range of a 9-point scale, and 4 were rated as "indifferent." All [illegible] tures had very small [illegible]

[illegible] in Task 1 [illegible] pairings in Task 2, such that each syllable was paired with both

ber was paired with both a P- and an I-paired [illegible] used for both lists, with the restriction that pleasant and indifferent items appear in first [illegible] quentially assigned to each order for their first trial.

Table 1 shows that for both male and female Ss [illegible] correct to the I-paired syllables when they were in [illegible] the stimulus position. Pooling the data for male and female Ss produced a significant Pleasantness × Position of Syllables interaction, $F(1, 76) = 5.63$, $p < .05$. The superior performance of male Ss to female Ss was also significant, $F(1, 76) = 4.41$, [illegible]. Correlated means comparisons of the P- and I-paired syllables for the various groups [illegible] when in the response position, $t(24) = 1.89$, $p < .10$, for males, but not for females. The indifferent-pleasant difference did not approach significance for either of the stimulus position groups.

No significant influence of acquired pleasantness was found in the analyses of the response learning [illegible] effect or in interaction with other variables (probably [illegible] of P-paired syllables). The P-paired syllables yielded higher pleasan[illegible] I-paired syllables for all 4 groups, $F(1, 76) = 6.96$, $p < .01$, [illegible] superiority being equally great in the stimulus and the response groups.

Experiment II. The superiority of P- to I-paired syllables found in Experiment I for the response position was a weak effect. However, the pairing of [illegible] bles with the pictures in the first task was simultaneous rather than successive as in previous experiments.[2] Since the acquired pleasantness effect is presumed to be the consequence of classically conditioned affective responses to the syllables [illegible] beneficiary of acquired pleasantness can be advanced with certainty only with data from a task that [illegible] completely than an anticipation task can [illegible] either by response recall (in any order) or by associa-

TABLE [illegible]

Measure	Position of syllable			
	Response		Stimulus	
	P	I	P	I
Female Ss				
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Male Ss				
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Note.—Abbreviations: P = syllable paired with pleasant picture, I = syllable paired with indifferent picture.

tive matching, both written (the former test being [illegible] sensitive to response learning, while the [illegible] only by associative matching and were included to control for any possible effect on similarity between test and study trials.

A total of 72 male Ss were assigned in randomized blocks to the 6 groups. The apparatus, lists, and first task were identical to those of Experiment I M, except that the number of trials was increased to 9 in order to strengthen the conditioning of positive affect to the syllables.

In the groups given anticipation training in the second task, S was shown the stimulus term alone for 3 sec., followed by the correct pair for 3 sec., with no overt response required. In the groups given associative-match training, S was shown the stimulus term along with all 8 response terms for 3 sec., followed by the correct pair for 3 sec., with [illegible] required. In both types of study [illegible] erent random order of the pai[illegible] between study and test trials. Following the second task, all Ss rated the nonsense syllables for pleasantness on a 7-point scale.

Table 2 shows the mean number of correct responses for both the first test trial and all 4 test trials, as well as the mean pleasantness ratings. The only significant differences between P- and I-paired syllables were found when the syllables were in the response positions and when Ss had been given anticipation training. A surprising result was that the P-paired syllables were superior as response

TABLE 2

MEANS OF MAJOR STATISTICS IN EXPERIMENT II

Measure	Training-testing sequence: Anticipation–recall		Anticipation–matching		Matching matching	
	P	I	P	I	P	I
Syllable in response position						
M correct responses						
Trial 1	2.08*	1.17	1.92	1.67	2.08	2.33
All trials	10.83**	8.42	11.83*	10.00	12.67	12.33
Pleasantness rating	5.05*	4.10	4.63*	3.89	5.17*	4.03
Syllable in stimulus position						
M correct responses						
Trial 1	1.50	1.33	1.83	2.00	2.42	2.25
All trials	11.00	11.75	11.92	11.92	12.83	11.50
Pleasantness rating	4.98*	3.76	4.83*	3.97	5.37*	4.11

Note. Abbreviations: P = syllables paired with pleasant pictures, I = syllables paired with indifferent pictures.
* $p < .05$.
** $p < .01$.

terms to the I-paired syllables even by the associative matching test, provided that anticipation training had been used. This superiority, however, was clearly smaller and appeared later in training than that obtained by the response recall test (the latter was largest on Trial 1). This pattern is consistent with the typical finding that the primary locus of indifferent–pleasant differences is prior to the first response, whether it is correct or not. As in Experiment I, the P-paired syllables were rated as more pleasant than the I-paired syllables regardless of whether or not PA performance was superior with them.

Discussion. Both of the experiments reviewed here confirmed the prediction that acquired pleasantness affects PA learning when it is manipulated on the response side, but not when it is manipulated on the stimulus side. The expectation that the acquired-pleasantness effect is essentially a response effect was based on the hypothesis that the primary mechanism for the effect is a disposition to select P-paired items for earlier and more vigorous response rehearsal (Silverstein, 1972a). This hypothesis received additional support in Experiment II, in which response recall was found to benefit from acquired pleasantness to a greater extent and earlier in training than did associative matching. The small superiority found in the associative-matching test for P-paired responses suggests a secondary mechanism for the acquired-pleasantness effect that operates exclusively within the association-learning stage. Silverstein (1972a) hypothesized such a secondary mechanism after finding a small superiority for P-paired responses in *homogeneous* (with regard to pleasantness) test lists, located exclusively within the association-learning stage, and suggested that it was based upon P-paired items having acquired a more reliable capacity to evoke distinctive and effective mediators. But even this association-learning effect requires active response rehearsal, since it appeared only after anticipation training and only on the response side of PA learning.

The fact that the P-paired syllables were rated significantly higher in pleasantness than the I-paired syllables in all conditions of both experiments demonstrates that the absence of superior performance for P-paired syllables in some experimental conditions was *not* the result of their having failed to acquire affective characteristics. Silverstein (1973) has shown that there is often no correlation between judgments of affective quality and performance in learning tasks.

REFERENCES

ANISFELD, M., & LAMBERT, W. E. When are pleasant words learned faster than unpleasant words? *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 132–141.

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs* 1960, 74 (10, Whole No. 497).

BATTIG, W. F., & SPERA, A. Rated association values of numbers from 0–100. *Journal of Verbal Learning and Verbal Behavior*, 1962, 1, 200–202.

SILVERSTEIN, A. Acquired word value, verbal learning and retention: Replication and extension. *Psychonomic [illegible]*, 1966, 18, 357–358.

SILVERSTEIN, A. Acquired pleasantness and [illegible]ate learning in mixed and homogeneous lists. *Journ[illegible] Experimental Psychology*, 1972, 93, 111–117.(a)

SILVERSTEIN, A. The role of task-anxiety in removing the effects of acquired pleasantness in paired-associate learning. *Journal of Experimental Psychology*, 1972, 94, 173–178.(b)

SILVERSTEIN, A. Acquired pleasantness and conditioned incentives in verbal learning. In D. E. Berlyne & K. B. Madsen (Eds.), *Pleasure, reward, preference*. New York: Academic Press, 1973.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. New York: Lippincott, 1960.

(Received April 9, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 537–539

SUBJECTIVE CORRELATION AND THE SIZE NUMEROSITY ILLUSION [1]

MICHAEL H. BIRNBAUM [2]
University of California, San Diego

MARC KOBERNICK AND CLAIRICE T. VEIT
University of California, Los Angeles

Judgments of the numerousness of patterns of dots vary with the size of the background on which the dots are presented. The direction of this judgmental illusion depends upon the experimental correlation between background size and actual number of dots. When the correlation is positive, judged numerousness varies inversely with background size; however, when the correlation is negative, the illusion is reversed. The subjective correlations estimated from an expectancy-contrast model were found to be a monotonic function of the 7 actual correlations used in 2 experiments. These results replicate and extend previous findings for the size–numerosity illusion.

An expectancy-contrast model for the size–numerosity illusion was proposed and tested by Birnbaum and Veit (1973). According to the model, the judgment of numerousness reflects an additive contrast between the numerousness of the dots and an expectancy produced by the context. The expectancy is assumed to depend upon the subjective correlation between numerousness and background size. In simplified form, the model can be written:

$$J = N - R_{NB}B,$$

where J is the impression of numerousness, N is the numerousness of the dots apart from the background effect, R_{NB} is the subjective correlation between numerousness and background size, and B is the subjective size of the background. The product, $R_{NB}B$, is interpreted as the expectancy for numerousness, based on background size. Overt ratings of numerousness are assumed to be a linear function of the impressions.

Several implications of the expectancy-contrast model have received tentative experimental support. First, it should be possible to reverse the direction of the illusion by changing the sign of the subjective correlation. Second, this interaction between background size and correlation is multiplicative and thus should be located in the bilinear component. Third, the effects of background size and numerosity should combine additively, independent of the correlation. Although the data appeared to be in general agreement with the first 2 predictions, a small but statistically significant discrepancy was obtained in 1 condition for the third prediction.

This report replicates the previous work, with an improved experimental design, to provide larger and cleaner effects. The second experiment employs more levels of correlation to examine the relationship between actual and subjective correlation and also explores the effect of changing the correlation.

METHOD

The general experimental procedure was similar to that of Birnbaum and Veit (1973), with several improvements. The main differences were as follows: (*a*) greater variation in background size relative to the variation in number of dots, (*b*) greater manipulation of the correlation, and (*c*) a longer series of warm-up trials.

The stimuli were patterns of black dots, 2.5 mm. in diameter, on square white cardboard back-

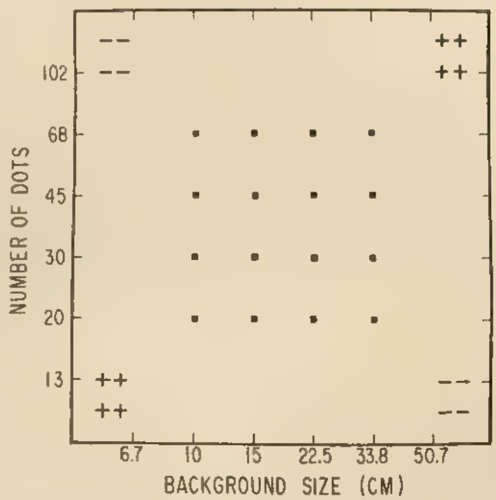


FIGURE 1. Diagram of the experimental design. (Solid squares represent test stimuli which were the same for all conditions. Plus signs represent contextual stimuli for positive correlations. Minus signs represent contextual stimuli for negative correlation conditions. Note that axes are spaced in logarithmic steps.)

[1] Computing funds for this research were supplied by the Campus Computing Network, University of California, Los Angeles. This paper was completed while the first author held a National Institute of Mental Health postdoctoral fellowship at the Center for Human Information Processing, University of California, San Diego, which provided additional support through National Institute of Mental Health Grant MH-15828. The authors thank Norman H. Anderson for comments on an earlier draft.

[2] Requests for reprints should be addressed to Michael H. Birnbaum, who is now at the Department of Psychology, Kansas State University, Manhattan, Kansas 66506.

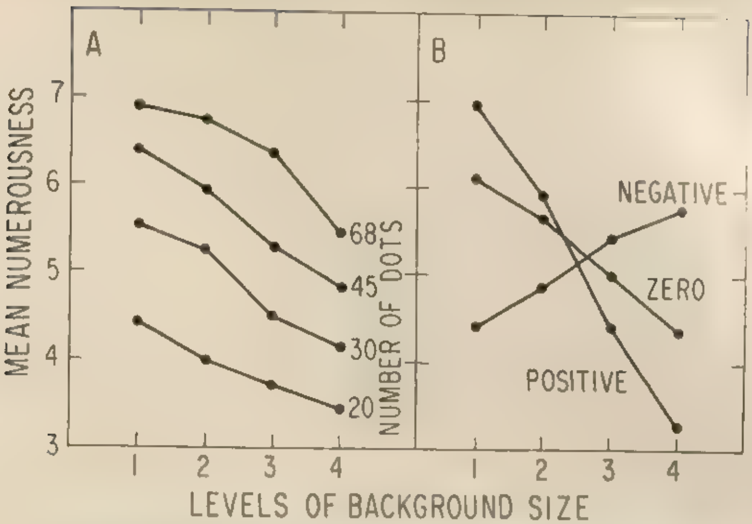


FIGURE 2. Panel A presents mean judgments of test stimuli, averaged across correlation conditions. Panel B, mean judgments of test stimuli, averaged across levels of number and plotted as a function of size for each correlation condition (Experiment I).

grounds of varying size. In presenting the stimuli, the blank side of the card was shown first, then the card was turned over so that *S* could see the dots for 5 sec. The *Ss* made their ratings on a 9-point scale, with 1 = very very few dots and 9 = very very many dots.

The experimental design is shown in Figure 1. The test stimuli (solid squares) represent a 4 × 4 (Number × Size) factorial design, in which either 20, 30, 45, or 68 dots were presented on square backgrounds with side lengths of either 10, 15, 22.5, or 33.8 cm.

The contextual stimuli were presented to manipulate the overall correlation between background size and number of dots. The contextual stimuli for the *positive correlation* (plus signs in Figure 1) paired 8–13 dots with background sizes of 5.0–6.5 cm., and 102–135 dots with background sizes of 54–60 cm.; thus, small numbers of dots appeared on small backgrounds and large numbers of dots appeared on large backgrounds. Contextual stimuli for the *negative correlation* (minus signs in Figure 1) paired 8–13 dots with background sizes of 54–60 cm., and 102–135 dots with background sizes of 5.0–6.5 cm., so that small numbers of dots appeared on large backgrounds and large numbers of dots on small backgrounds. The *zero correlation* condition included half of either type of contextual stimuli.

In Experiment I, the 21 University of California, Los Angeles, undergraduates were divided into 3 groups of 7 *Ss* each, with either positive, zero, or negative contextual stimuli for each group. Each group received an inducing series of 24 randomly ordered presentations of contextual stimuli. Following this series, there were 2 replicates of the 16 test stimuli randomly interspersed among 48 trials of contextual stimuli, giving a total of 104 trials.

In Experiment II, the ratio of contextual trials to test trials (and hence the correlation between size and number) was manipulated. The *Ss* were 30 University of California, Los Angeles, undergraduates divided into 6 groups of 5 *Ss* each. Each group received 16 presentations of contextual stimuli followed by either 36, 24, or 16 positive or negative contextual stimuli interspersed among the basic 16 test stimuli.

After the first series came the counterinducing series, in which 16 contextual stimuli of the opposite sign were presented, followed by the 16 test stimuli interspersed among 24 contextual stimuli of the opposite sign from the first series.

RESULTS

Experiment I. Figure 2A plots mean ratings of numerousness as a function of background size, with a separate curve for each level of numerosity, averaged over correlation conditions. According to the expectancy-contrast model (Equation 1), the curves in Figure 2A should be parallel, because the effects of background size and number of dots combine additively. Although the graphical appearance is somewhat rough, there do not appear to be any systematic deviations from parallelism. The parallelism prediction is supported by a nonsignificant overall Background × Numerosity interaction, $F\ (9,\ 162) = 1.42$. Analyzed separately for each correlation condition, this interaction was also nonsignificant, $Fs\ (9,\ 54) = 1.39$, 1.75, and 2.01 for the positive, zero, and negative conditions, respectively.

Figure 2B shows mean judgments of numerousness as a function of background size, with a separate curve for each correlation condition. Con-

sistent with theoretical predictions, judgments vary *inversely* with background size when the correlation is *positive* and *directly* when the correlation is *negative*. The illusion for the zero correlation group is intermediate, but similar in direction to the positive correlation, consistent with previous results (Birnbaum & Veit, 1973). The interaction between background size and correlation is statistically significant, F (6, 54) = 20.23, with about 98% of its variance in the bilinear component. The residual interaction is nonsignificant ($F < 1$) as were the Context × Number, F (6, 54) = 2.11, and the Context × Background × Number interactions, F (18, 162) = 1.67. These results replicate previous findings for the size-numerosity illusion (Birnbaum & Veit, 1973).

The expectancy-contrast model implies that the curves in Figure 2B should be linear functions of B with slopes proportional to R_{NB}. The curves are very nearly linear functions of the a priori levels of background size, which were chosen in equal log steps. The slopes (reversed in sign) are 3.7, 1.7, and −1.4 for the positive, zero, and negative correlations, respectively. The present illusory effects are much greater than those reported by Birnbaum and Veit (1973). The larger illusions are probably due in part to the improved experimental design and the greater manipulation of the experimental correlation. Part of this increase may be due to a range effect on the category rating scale, attributable to the smaller variation in actual number of dots relative to the variation in background size.

Experiment II. The values of R_{NB} for the first series, estimated from Equation 1, were 5.6, 3.9, 2.8, −1.1, −1.5, and −1.8 for the 6 conditions, in order of decreasing actual correlation. The trend is clearly monotonic. This finding lends further support to the notion of subjective correlation by showing that R_{NB} is sensitive to the actual correlation employed.

The estimated values of R_{NB} for the second series of Experiment II showed a small and nonsignificant effect of the correlation used in the first series; they were positive when the second correlation was positive and negative when the second correlation was negative. The intervening series of 16 reversed contextual trials may have been sufficient to largely erase the preceding experience.

Discussion. The fact that the illusion can be manipulated and reversed through manipulation of the correlation supports the expectancy interpretation of the illusion. The slope for the zero correlation condition would be interpreted to imply that everyday experience produces a positive correlation between size and number (Birnbaum & Veit, 1973). By analogy, these results support an expectancy account of the size-weight illusion (Anderson, 1970).

These results replicate and extend the previous findings for the size numerosity illusion. The improved experimental design appears to produce larger and cleaner effects (Experiment I). The fact that subjective correlation estimated from the model is a monotonic function of actual correlation (Experiment II) provides further support for the expectancy-contrast model of Birnbaum and Veit (1973).

REFERENCES

Anderson, N. H. Averaging model applied to the size-weight illusion. *Perception & Psychophysics*, 1970, 8, 1-4.
Birnbaum, M. H., & Veit, C. T. Judgmental illusion produced by contrast with expectancy. *Perception & Psychophysics*, 1973, 13, 149-152.

(Received April 14, 1974)

Journal of Experimental Psychology
1974, Vol. 102, No. 3, 539-542

SPATIAL STIMULUS GENERALIZATION AS A FUNCTION OF WHITE NOISE AND ACTIVATION LEVEL

ROBERT E. THAYER[1] AND DAVID CAREY

California State University, Long Beach

Three experiments (91, [illegible], and 68 Ss) were conducted to study spatial stimulus generalization with voluntary S response as a function of white noise and activation level. Activation was assessed by several verbal report measures. Stimulus generalization consistently [illegible]creased with noise, but changes were apparently unrelated to measured activation level. The results do not entirely rule out activation as a mediator, and if a relationship exists, it could involve more than 1 activation dimension. Arguments are made concerning use of the present experimental task and procedure to assess hypotheses concerning attentional selectivity.

Several studies have been conducted on spatial stimulus generalization with voluntary responses. Typically, Ss are trained to respond rapidly to a

[1] Requests for reprints should be sent to Robert E. Thayer, Department of Psychology, California State University, 6101 East Seventh Street, Long Beach, California 90840.

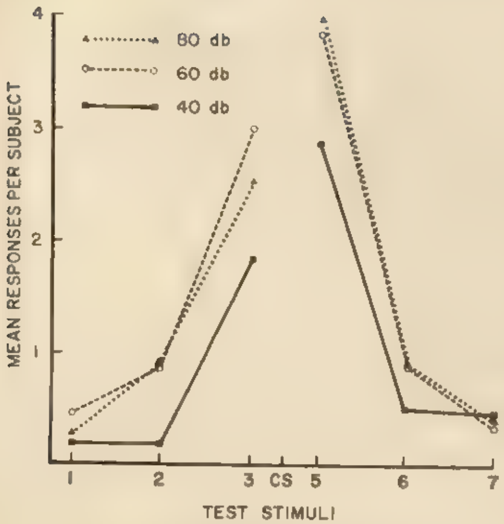


FIGURE 1. Spatial generalization gradients as a function of white noise.

center stimulus and are tested on successive trials when peripheral stimuli are occasionally presented. Generalization gradients as a function of stimulus proximity are then observed.

Previous research has focused on generalization as a function of such variables as "speed-up" instructions, drive, anxiety, and shock (Harleston, 1961; Mednick & Freedman, 1960). Activation is probably common to all these variables. (See Thayer & Moore, 1972, for a further discussion of this point.) However, if it is to be argued that activation is a mediating variable, it is important to provide some kind of direct arousal assessment. Otherwise, diverse underlying physiological states might be confused with a single activation dimension.

While physiological activation measurement is quite complicated, the development of the Activation Deactivation Adjective Check List (AD ACL; Thayer, 1967) offers an easy way of determining multidimensional activation changes from experimental manipulations. The AD ACL is composed of a group of activation-descriptive adjectives which *S*s use to rate their momentary activation feelings. The adjectives have been factor analyzed into 4 independent dimensions: General Activation (e.g., lively, active, full of pep), High Activation (e.g., clutched up, jittery, stirred up), General Deactivation (e.g., at rest, still, leisurely), Deactivation-Sleep (e.g., sleepy, tired, drowsy). These activation dimensions were shown to substantially correlate with physiological measurements and to be sensitive to a variety of activation-inducing conditions. (See Thayer & Moore, 1972, for a current review of validation research.)

The purpose of the present research was twofold. The first purpose was to study spatial stimulus generalization with voluntary *S* response as a function of activation and to assess activation directly with the AD ACL. A second research purpose was to study the effects of white noise as an influe... on generalization. Noise is associated with acti... effects, but previously has not been show... crease generalization.

Experiment I

Method. The *S*s were 91 male and fema... ductory psychology students. The appara... cluded 4 visual displays constructed in a ... pattern, thus accommodating 4 *S*s at a ru... display board was straight and included a ho... row of 7 white 6-v. lamps, 1½ in. apart, and ... red 6-v. warning lamp 2⅝ in. above the cen... lamp. The *S*s sat in fixed chairs so their ... approximately 46-in. from the center w... Responses to the stimulus lights were ... presses of a noiseless button. Either 40, 60, or 80 db. of white noise was presented by a Gras... Stadler Model 901B white noise generator and 5-in. speaker equidistant from each *S*. The experiment was conducted in a soundproof room.

Using a between-*S*s design, 31 *S*s (balanced for sex) were tested in 40- and 60-db. conditions, and 29 *S*s were tested in the 80-db. condition. The *S*s were given 5 practice trials and 25 conditioning trials, during which the center lamp was flashed every 10 sec. (.3-sec. duration). The test trials followed immediately and included 7 presentations of each peripheral lamp and 10 booster presentations of the center lamp. Five seconds before every conditioning and test presentation, a w... flashed (.3-sec. duration). The ... center presentations were distributed th... 52 test trials in 1 of 3 random orders. The A... was given to each *S* at the end of the test serie...

The *S*s were told that this was a reaction ... experiment, and that they must press the button ... fast as possible whenever the center lamp was lit. They were also told not to be concerned about errors, but to try to press only when the center lamp was lit.

TABLE 1

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) OF PERIPHERAL ERRORS (PEs) AND CENTER RESPONSES (CRs) AS A FUNCTION OF WHITE NOISE

Experiment	Noise level and response measures					
	40–45 db.		60 db.		80 db.	
	PE	CR	PE	CR	PE	CR
I	6.1 (3.9)	22.7 (5.3)	9.5 (5.0)	26.8 (5.4)	9.0 (4.1)	25.3 (5.7)
II	5.7 (3.4)	28.1 (5.3)			6.6 (4.0)	28.8 (4.5)
III	4.8 (3.4)	29.0 (4.5)			7.9 (6.6)	30.1 (3.8)

Note. Half of the center responses in Experiment II were multiplied by 1.66 to equate frequency of conditioning trials across counterbalanced conditions.

[2] In this and the 2 subsequent experiments, noise was presented continuously throughout conditioning, testing, and AD ACL administration.

Results. The response measures were defined as follows: *Peripheral errors* occurred during the test period when S responded to a peripheral light as though it were a center light; *center responses* were correct responses to the center light during the conditioning *and* test periods.

Figure 1 indicates the shape of the generalization gradients for the 3 noise conditions. It can be seen that gradients for 40 db. are flatter than those for 60 and 80 db., particularly from Test Stimuli 1, 2, 6, and 7.

Table 1 presents the means and standard deviations of peripheral errors and center responses in each of the noise conditions. Analysis of variance in peripheral errors showed the expected significant effect of intensity, $F\ (2, 88) = 5.34$, $p < .05$. Post comparison tests showed significant differences in the 40:80 and 40:60 db. comparisons, ts $(58, 60) =$ 2.79 and 2.94.

While the above results indicate increased generalization as a function of noise, changes might be due to greater responsiveness to all stimuli, including center stimuli, or primarily to peripheral-stimuli responsiveness. Analyses of center responses showed no significant differences, though the responses did increase as a function of noise. A comparison of the strength of effects (omega squared) for the 40:80 db. conditions showed that noise accounted for 10% of the peripheral-error variation, and 4% of the center-response variation.

The anticipated reported-activation differences as a function of noise were not obtained, nor was there a correlation between errors and activation. The differences across noise conditions were small and inconsistent. There was some tendency for the AD ACL General Activation subtest to be related to peripheral errors within each condition, but the relationships were weak.

Experiment II

Since Experiment I showed noise effects on generalization, but highly variable activation reports, an attempt was made to further investigate the activation relationship with somewhat less error variance. Thus, the manipulations of Experiment I were repeated in a within-*S*s design.

Method. Forty-two male and female *S*s from introductory psychology classes were run in a counterbalanced design (half received 45 db. first and half, 80 db. first). The apparatus and materials were the same as those described in Experiment I, except that tests were not conducted in a soundproof room.

The *S*s were again given 25 conditioning trials and 52 test trials, followed by the AD ACL. They then received 15 more conditioning trials, a second set of 52 test trials, and another AD ACL. An attempt was made to duplicate most of the conditions in Experiment I; however, 45 db. was chosen instead of 40 db. because of a higher uncontrollable ambient noise level in the test room.

Results. Consistent with Experiment I, peripheral errors increased with noise (Table 1), $t\ (41) = 1.69$, $p < .10$. Center responses also increased slightly. Because of the within-*S*s design employed, it is somewhat difficult to estimate strength of independent variable effects (omega squared), but the most conservative estimate would be that noise accounted for 2% of peripheral-error variation, and essentially zero center-response variation.

The AD ACL subtests General Activation and Deactivation-Sleep apparently were not influenced by noise, but substantial noise effects were observed for the High Activation and General Deactivation subtests, $t\ (41) = 2.53$, $p < .02$; $t\ (41) = 4.42$, $p < .001$. Although the activation measures indicated a noise effect, the activation change scores were not correlated with the perceptual-performance change scores.

Experiment III

Since the generalization differences in Experiments I and II apparently were not related to reported activation, the present experiment was designed to investigate alternative individual-difference measures. To investigate the possibility that noise was creating activationlike changes not sensed by the AD ACL, 2 anxiety trait measures were included. Also, *S*s were asked to rate and verbally describe their reactions to the noise.

Method. As in Experiment I, the experimental design was a between-*S*s type; it included one 45- and one 80-db. condition. Sixty-eight male and female introductory psychology students were used. The apparatus and procedure were the same as in Experiment I, except that *S*s were again tested in a nonsoundproof room.

In addition to the AD ACL, the present experiment employed the Manifest Anxiety Scale, a shortened version of the S-R Inventory of Anxiousness, and 2 5-point rating scales on which *S*s could rate the extent that noise generally disturbed their concentration and, specifically, the degree to which the experimental noise influenced their performance. Finally, one question requested a paragraph description of how the experimental noise disturbed performance.

Results. Similarly to Experiments I and II, high noise resulted in increased peripheral errors (Table 1), $t\ (66) = 2.42$, $p < .01$. Center responses also increased, but not to a statistically significant degree. The omega-squared index indicated that 7% of the variation in peripheral errors could be accounted for by noise, while less than 1% of the center-response variation was due to noise.

The AD ACL High Activation subtest may have been influenced by noise, $t\ (66) = 1.69$, $p < .1$, but the other subtests showed small changes. Again, none of the AD ACL scores was reliably related to peripheral errors or center responses.

Ratings by *S*s indicated greater interference from the 80- than the 45-db. condition, $t\ (66) = 2.80$, $p < .01$, but these ratings were not correlated with errors. The other individual-difference measures did not predict peripheral errors or center responses in any reliable and consistent way. Only 56% of the *S*s in the 80-db. condition indicated that noise

interfered with their reactions to the light. Although there was no obvious consistency in *Ss'* verbal explanations of how the sound interfered, one provocative note is that several *Ss* said the noise interfered by making them tense and nervous, while several others said it interfered by making them sleepy and drowsy.

Discussion. Consistent results from all 3 experiments clearly demonstrate that loud noise (i.e., 80 db.) increases spatial stimulus generalization. Furthermore, this effect is primarily due to increased peripheral errors, not increased general responsiveness. Comparisons of the strength of effects of noise between peripheral errors (incorrect responses) and center responses showed that the peripheral-error changes were substantially greater than center-response changes in all 3 experiments. The influence of noise on generalization seems to be the same as special instructions continually prompting *Ss* to speed their responses (Harleston, 1961), and it may be equivalent to such variables as anxiety and drive (Mednick & Freedman, 1960).

However, contrary to expectations, stimulus generalization was not obviously related to [illegible] activation. Although the AD ACL has been shown to substantially correlate with psychophysiological measures and to reliably predict activation-inducing independent variables (Thayer, 1967; Thayer & Moore, 1972), the activation measures did not consistently predict the observed perceptual-performance effects. This is true even though certain AD ACL subscales apparently monitored noise effects reliably in one experiment. Other activation-related individual difference measures, particularly 2 measures of anxiety, did not predict generalization effects either.

Nevertheless, the similarity of the present noise-[illegible] generalization effects to previously demonstrated effects caused by speed-up instructions suggests the common construct of activation. Several [illegible]

states has been demonstrated elsewhere [illegible] Moore, 1972), and it creates substantial [illegible]tions in predicting behavior without [illegible]priate individual-difference measure[illegible]

The present experimental task and [illegible] an interesting possibility for testing [illegible] noise and activation level on attentional s[illegible] Broadbent (1971) has proposed that [illegible] arousal result in the focusing of attenti[illegible] peripheral to central cues when peripheral [illegible] unimportant.

In order to demonstrate selectivity in th[illegible] results, it would be necessary to show [illegible] responsiveness to the center stimulus and [illegible]duced responsiveness to peripheral lights under [illegible] activation conditions. Although loud noise [illegible] produce more responses to the center stimu[illegible] all 3 experiments, there was not reduced responsive[illegible]ness to peripheral cues; rather, as described abo[illegible] [illegible] greater degree than center responses. The presen[illegible] [illegible] Broadbent's (1971) expectations. However, 2 points lead to caution in direct application of these results. First, the nature of the present experimental task [illegible] have introduced confounding psychomotor performance effects in addition to the expected percept[illegible] attention effects. Second, Broadbent and his associates have used higher noise levels to produce arousal than those employed in the present research

* The authors wish to thank [illegible] concerning the relationship [illegible] and attentional selectivity.

REFERENCES

Broadbent, D. E. *Decision and stress.* London: Ac[illegible] Press, 1971.

Harleston, B. W. The effects of speed-up instructions [illegible]

Mednick, S. A., & Freedman, J. A. Stimulus generalization. *Psychological Bulletin,* 1960, 57, 169-200.

Thayer, R. E. Measurement of activation through self-report. *Psychological Reports,* 1967, 20, 663-678.

[illegible]

(Received April 21, 1975)

JOURNAL OF EXPERIMENTAL PSYCHOLOGY MONOGRAPH

Vol. 102, No. 3, 543-561 March 1974

THE NONADDITIVITY OF PERSONALITY IMPRESSIONS

MICHAEL H. BIRNBAUM[1]

University of California, Los Angeles

Ratings of the likableness of persons described by 2 adjectives showed con- [illegible] effect of either adjective varied directly with the likableness of the other [illegible] theoretical question of whether interactions were due to nonlinearity in the [illegible] inconsistent with additive models, confirming the interpretation that the interactions are "real" and should not be scaled away. Theoretical and methodological implications are discussed.

[illegible]

fulfillment of the requirements for the PhD [illegible] from [illegible]

[illegible] overall impression [illegible] would be described by both adjectives, and rate how much he would like such a person.

Figure 1 provides a schema for the analysis of impression formation. The 2 stimulus adjectives are referenced by the indices, i and j. The psychological representations of the adjectives, s_i and s_j, are combined by the integration function, I, to form the psychological impression, Ψ_{ij}, which is transformed by the response function, J, to the overt response, R_{ij}.

There are 3 problems to be solved: (*a*) finding the appropriate stimulus representation (s), (*b*) determining the integration function (I), and (*c*) finding the

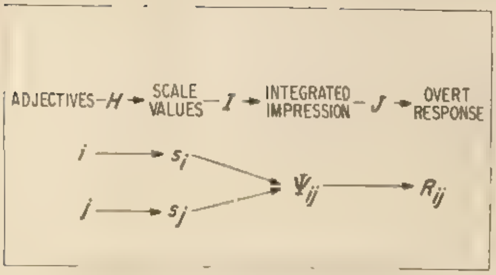


FIGURE 1. Schema for discussing research on information integration.

relationship (J) between the overt responses and the integrated impressions. Any theory of impression formation must deal, at least implicitly, with all 3 of these problems.

ADDITIVE AND CONSTANT-WEIGHT AVERAGING MODELS

A simple linear model provides an example of such a theory. The adjectives are assumed to be represented as single values on a likableness continuum. The I function, defined on all adjective combinations, would be a linear combination of the values of the adjectives:

$$\Psi_{ij} = w_0 s_0 + w_1 s_i + w_2 s_j, \qquad [1]$$

where w_1 and w_2 are the weights reflecting position in the set, s_i and s_j are the scale values of the adjectives, and $w_0 s_0$ is the postulated initial impression, reflecting what the impression would be in the absence of information.

Anderson (1962) pointed out that if sets of adjectives were constructed according to factorial designs, then the analysis of variance test for interactions provides a powerful test of Equation 1. If the adjectives were to change their meaning nonlinearly in combination, if the I function was nonadditive, or if the judgment function (J) was nonlinear, then significant interactions would be expected. If they were nonsignificant, there would seem to be no reason to postulate such complications. The data for the majority of Anderson's (1962b) 12 *S*s appeared to be in rough agreement with Equation 1.

The *constant-weight averaging model* (Anderson, 1968b) is a special case of Equation 1 that requires the weights to sum to 1; the *additive* model places no restriction on the sum of the weights. Critical tests of averaging vs. adding models (in which the number of items in the set are varied) favor the averaging formulation over the additive (Anderson, 1971, 1974, in press). However, when the number of items in the set is held constant, as in the present experiments, the models are equivalent. Therefore, the present article uses the term *additive* to include both additive and constant-weight averaging models.

Anderson's (1968b) review of the evidence for the constant-weight averaging model concluded that

> The model always fits the data quite well, but there are almost always small, significant discrepancies. Inspection of the data has failed to reveal the origin of the discrepancies; they may reflect some fundamental error in the model, or they may result from remaining shortcomings in the experimental technique [p. 736].

Recent studies have shown large discrepancies from the constant-weight averaging model for several information integration tasks. Interactive models have been proposed to account for these nonadditivities (Anderson, 1972; Birnbaum, 1972a, 1972b, 1973; Birnbaum, Parducci, & Gifford, 1971; Lampel & Anderson, 1968). However, there are 2 possible interpretations of the discrepancies that have been observed in these studies: (*a*) the integration (I) of information is not an additive or simple averaging process, or (*b*) the judgment function (J) is nonlinear. The next section shows how previous conceptualizations cannot differentiate these alternatives.

MODEL TESTING AND MEASUREMENT

Assuming Validity of Responses

When the J function is assumed to be linear, models of impression formation can be tested by comparison of theoretical predictions with the raw data. Functional measurement (Anderson, 1970) finds the

stimulus representation in accord with the model to be tested. These stimulus parameters are then used as the basis for statistical tests of fit. If either the model *or* the response scale was incorrect, it would show up as a significant discrepancy. When the data fit the model, the fit is usually interpreted as joint support for both the model and the response scale.

Ordinal Tests

When the J function is only assumed to be strictly monotonic, it becomes more difficult to discriminate among different models. Conjoint measurement analysis (Krantz & Tversky, 1971) describes conditions that ideal data would have to satisfy to be *ordinally* consistent with the theory. For example, crossover interactions would be ordinally inconsistent with additive models, for no monotonic transformation could make the data fit the model. In this case, everyone agrees that the model should be rejected. It should be noted that ordinal violations of additivity will show up as significant interactions in the analysis of variance. The problem arises when significant interactions occur in the *absence* of ordinal violations.

Assuming Validity of the Model

When the additive model is assumed to be valid, then the analysis of variance tests the linearity of the J function. A nonlinear J function will produce significant interactions even though the underlying integration is additive. In the absence of ordinal violations of the model, it is possible to find a monotonic transformation, J^{-1}, which rescales the data to additivity.

When Both Model and Response Scale are in Doubt

Krantz, Luce, Suppes, and Tversky (1971, p. 445) have taken the extreme view that when discrepancies from additivity can be removed by monotonic transformation, they should be attributed to nonlinearity in the response scale, rather than to nonadditivity of the integration function. However, when a monotonic transformation would bring otherwise contradictory data into line with the model, the status of the model remains uncertain. This procedure *assumes* the validity of the model; hence, the existence of such a transformation does not mean that the model is validated. To resolve this difficulty, it is necessary to place transformations within the scope of psychological theory and to provide additional constraints which determine their appropriate application.

Advances in Model Testing

These 4 experiments provide a progressive sequence that systematically eliminates the additive or constant-weight averaging models. The experiments illustrate novel approaches to model testing that remove the difficulty of deciding whether or not to rescale the data. Since these techniques will be of interest to psychologists in many areas, they are outlined briefly below.

The first 2 experiments apply criteria of stimulus and response scale consistency. Experiment I assumed the validity of a priori values for the adjectives obtained in previous work (Anderson, 1968a) and required that the integration model yield scale values that are linearly related. Experiment II investigated the effects of different response procedures, thought to produce different J functions. By a principle of convergent operationism (Garner, Hake, & Eriksen, 1956), if similar interactions are obtained with different response procedures (operational definitions of Ψ), then the interpretation that the interactions are "perceptual" is enhanced.

Experiment III illustrates how the simultaneous evaluation of 2 or more integration processes can provide the leverage to define the concept of an *appropriate* transformation (Birnbaum, 1972b; Birnbaum & Veit, 1974). *Stimulus scale invariance* requires that the scale values (s in Figure 1) be independent of task. *Response scale invariance* requires that the J function be independent of task.

In Experiment III, *S*s performed 2 in-

tegration tasks, rating the difference in likableness between the 2 adjectives as well as the likableness of a person described by the combination. Since the same stimuli and response scale were employed for both tasks, response scale invariance requires that the *same* J-inverse transformation be applied to both tasks. The stimulus scale convergence criterion defines an *appropriate* rescaling of the combination ratings as one which *both* makes an hypothesized model fit the data *and* leads to the derivation of scale values that agree with those derived from the difference task.

Although rescaling of the data might make the model fit a single set of data, the transformation that reduces the interactions implies which psychological differences are greater than or equal to others. Experiment IV obtained direct ratings of these differences, and it tested whether these ratings are qualitatively consistent with the transformation that makes the data additive.

Experiment IV required only the ordinal information in the data to reject the additive and constant-weight averaging models. This leverage was provided by a compound integration task in which *S*s rated the difference in likableness between pairs of hypothetical persons, each described by a pair of adjectives. The ratings can be rescaled to fit the subtractive model; this rescaling may or may not make the additive model of impression formation fit. Thus, difference judgments can be used as a basis for response rescaling to provide a scale-free test of the constant-weight averaging model of impression formation.

Experiment I: Test of Additive Models

The first experiment was designed to uncover the nature of the discrepancies previously observed. Although the basic procedures were similar to those of the research described by Anderson (1968b), these were modified to permit a clearer assessment of the expected interactions. Thus, only 2 factors were employed, but each was represented by a greater number of levels covering a wider range. This permitted a greater variation of within-set range.

Method

The *S*s were presented with pairs of personality-trait adjectives and were instructed "your task is to imagine a person who would be described by *both* of the traits and judge how much you would like such a person." The *S*s recorded their judgments in numerical form, using 1 of 9 ratings for each pair: 1 = dislike very very much, 2 = dislike very much, 3 = dislike, 4 = dislike slightly, 5 = neutral (neither like nor dislike), 6 = like slightly, 7 = like, 8 = like very much, and 9 = like very very much.

Subjects. The *S*s were 300 University of California, Los Angeles, undergraduates fulfilling a requirement in introductory psychology. One hundred different *S*s served in each replicate of the experiment.

Stimuli. The adjectives were taken from Anderson's (1968a) list of 555 common personality-trait adjectives. Each stimulus replicate consisted of a set of 25 adjective pairs produced from a 5 × 5 (Adjective A × Adjective B) factorial design. The 5 levels of likableness of each adjective factor were separated by steps of approximately 1.28 on Anderson's 0–6 normative scale. The adjectives are printed (with normative values in parentheses) in the margins of Figure 2.

Procedure. The 25 adjective pairs were printed in 1 of 6 random orders on the same page with the instructions and the labeled response scale. The adjectives in each pair appeared side by side, with the adjective from Factor A on the left for half of the forms. The *S*s were instructed to read through the entire list before beginning to record their ratings.

Results and Discussion

Additive and constant-weight averaging models. The 3 panels of Figure 2 show the mean judgment of each adjective pair, averaged over all *S*s within the replicate. The mean judgments are plotted as a function of the normative value for the adjective from Factor A, with a separate curve for each adjective from Factor B. The slope of the curves represents the effects of the Factor A adjective. The vertical differences between the curves represent the effects of the Factor B adjective. According to additive or constant-weight averaging models, the curves in each panel should be parallel; that is, the effect of one adjective should not

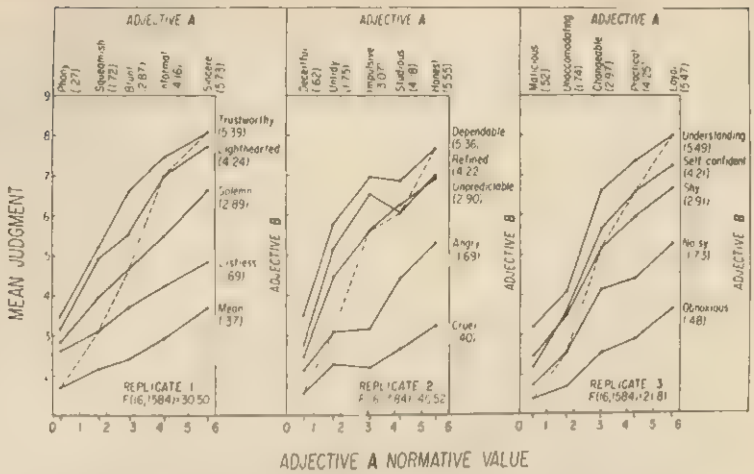


FIGURE 2. Mean ratings of likableness of pairs of adjectives: Experiment I. (Anderson's, 1968a, normative values for individual adjectives are listed above and to the right of each panel.)

depend upon the particular adjective with which it is paired. Instead, the curves in each panel diverge to the right; the effect of either adjective appears to be proportional to the likableness of the other adjective describing the same person.

This divergence was characteristic of the data for most of the individual *S*s: 88%, 88%, and 89% of the *S*s in the respective replicates showed this form of interaction. In Experiments I–IV between 85–95% of the single *S*s resembled the group data. The A × B interactions are highly significant for each of the replicates; *F* values for these interactions are indicated in the panels of Figure 2. Of course, the *F* is much larger when the data from all 3 replicates are combined, F (16, 4752) = 72.80. If the response scale of this experiment is assumed to be valid, the interactions shown in Figure 2 can be interpreted as evidence against the additivity of impressions.

Although the A × B interaction varies between stimulus replicates, F (32, 4752) = 8.28, this higher order interaction is small in comparison with the similarity of the nonadditivities. There are, however, a few peculiarities that appear to depend upon the meanings of the particular adjectives. For example, SELF-CONFIDENT & MALICIOUS is less likable than SHY & MALICIOUS although SELF-CONFIDENT is more likable than SHY in combination with other traits. A SELF-CONFIDENT, malicious person may be perceived as more likely to carry out malicious actions than a SHY one. But the divergent interaction occurs in all 3 replicates and represents the main source of difficulty for the additive models.

Multiplicative model. The multiplicative model, $\Psi_{ij} = s_i s_j$, predicts that the curves in each panel should be a diverging fan of linear functions. The interaction should be located entirely in the bilinear component (Anderson, 1970). Although the major portion of the variance of the interaction (81.6%, 56.2%, and 86.9% for the 3 replicates) was located in the bilinear component, the residuals were also statistically significant, Fs (15, 1485) = 6.36, 20.68, and 3.26. The nature of the discrepancy can be seen most clearly in the second panel of Figure 2. According to the multiplicative model, the curves should be linear; instead, the upper curves are negatively accelerated relative to the lower curves. The other 2 replicates show the same discrepancy, but to a lesser degree.

Range model. The range model was fit to the data (averaged over replications) following the general procedures outlined in Birnbaum et al. (1971), and using

Anderson's (1968a) normative values as estimates of the scale values. The range model, using just one additional parameter, accounts for some 80% of the variance left unaccounted for by the constant-weight averaging model.

Averaging model with differential weights. The mean absolute discrepancy was only .05, indicating a very good fit to 15 points using 10 parameters. The best-fit estimates of scale values for the 5 levels were 1.49, 3.02, 5.24, 6.48, and 7.83; the best-fit weights were 1.92, 1.18, .67, .73, and .90. Fewer parameters are required with the assumption of the validity of Anderson's (1968a) values and by estimating the weights as a polynomial function of scale value. Although this procedure uses more parameters than the range model, the fit was not as good.

Rescaling. Interactions in the analysis of variance might be due to a nonlinear relationship between the impressions of likableness and the overt responses of the *S*s. Thus, the actual impressions may be additive, but the overt responses may be related only ordinally to the theoretically correct response scale. The possibility of response scaling was investigated using 2 techniques of data transformation.

The first technique assumes that Anderson's (1968a) normative values for the single items are appropriate estimates of the scale values and that judgments of pairs of items of equal value should be linearly related to the judgments of the single items (Birnbaum, 1972a). The transformation is then estimated as a polynomial by a least squares criterion. The advantage of this technique is that although it is capable of producing a radical rescaling, it will not eliminate "real" interactions that depend upon differences in within-set range, if the a priori scales are appropriate. As can be seen in Figure 2, the judgments of pairs of minimal within-set range (connected by dashed lines) are already nearly linearly related to normative ratings of single values. If anything, this function (treated in this procedure as the J function of Figure 1) is slightly negatively accelerated, so that the rescaling that makes it linear would actually increase the magnitude of the divergent interaction.

If the divergent interaction is real, then the distribution of psychological impressions is somewhat positively skewed (as given by the projection of the data points on the ordinate of each panel of Figure 2). Positive skewing would lead to a negatively accelerated judgment function according to range–frequency theory (Parducci & Perrett, 1971). This suggests that contextual effects operating on the J function would tend to counteract the effects of true interactions. Birnbaum et al. (1971, Experiment V), demonstrated that manipulation of the frequency distribution of physical means of sets of psychophysical stimuli influences the form of the interaction between the components in a manner predictable from range–frequency principles of judgment.

The second transformation procedure assumes that the integration is additive (or constant-weight averaging) and attempts to transform the data to fit the additive model (Kruskal, 1965). The monotone analysis of variance (MONANOVA) computer program (Kruskal & Carmone, 1969), applied to the mean judgments, greatly reduced the percentage of total variance in the interaction (from 5.0% to .4%). Thus, the data appear roughly consistent with the constant-weight averaging or additive model at an ordinal level. The fact that the data can be transformed to fit raises a difficult theoretical problem: Is the nonparallelism in Figure 2 due to nonadditive integration of information or to a nonlinear response function?

If the additive model is assumed to be correct, the MONANOVA analysis indicates that the judgments are a positively accelerated function of the impressions. Additionally, it would mean that the scale values for adjectives presented in pairs are a negatively accelerated function of Anderson's (1968a) values for the same adjectives presented singly. The positively accelerated function for J derived from the MONANOVA analysis can be interpreted by range–frequency theory (Parducci & Perrett, 1971) to indicate that

the psychological distribution of stimuli is negatively skewed. Kanouse and Hanson (1972) have also hypothesized that the distribution of evaluative stimuli is negatively skewed, but based their arguments on other considerations.

In short, Experiment I demonstrates that ratings of likableness are inconsistent with the additive models. However, 2 interpretations are consistent with the data. The first assumes that the ratings are valid measures of psychological impressions and concludes that the integration of information is nonadditive. The second assumes the validity of the constant-weight averaging (or additive) model and concludes that the responses are a positively accelerated function of the impressions. Both interpretations can account equally well for the data of Experiment I. Consequently, the following experiments were designed to discriminate them.

Experiment II: Variation of Response Scales

The earlier experiments yielding data more consistent with the additive models have used other procedures for obtaining *S*'s response (see Anderson, 1974, in press). The interactions observed in Experiment I may be due to nonlinearity in the *J* function that depends upon the particular procedure *S* uses to indicate his impressions. Therefore, Experiment II tested the additive models using several different procedures for responding to assess whether the interaction obtained in Experiment I is specific to the 9-category rating scale used in that experiment. Each of the 4 conditions of the present experiment tested a different interpretation of how the interactions might depend upon the method for responding: reversing the scale, endpoint anchoring with 20 categories, linemark responses, and matched pairs.

Although the matched pairs procedure has not been used before, it has the apparent advantage that it does not require a metric response from the *S*. Thus, this procedure avoids the argument that metric responding may induce *S*s to integrate separate implicit responses to the individual adjectives rather than forming an overall impression before responding.

Method

As in Experiment I, *S*s read pairs of personality-trait adjectives and judged how much they would like hypothetical persons who would be described by both adjectives. The conditions differed with respect to the instructions for the response. Each set of instructions was used in a separate small experiment, with different *S*s.

For the reversed scale, 1 = like very very much and 9 = dislike very very much. The *S*s were consequently instructed to rate how much they would *dislike* the hypothetical persons. The 20 *S*s rated the adjective pairs of both Replicates 1 and 2 of Experiment I.

The 20-category scale was anchored by instructions that 1 represented the likableness of a person who would be described by MEAN, PHONY, MALICIOUS, OBNOXIOUS, and LIAR; 20 represented the likableness of SINCERE, LOYAL, INTELLIGENT, UNDERSTANDING, and DEPENDABLE. The *S*s were instructed to judge each adjective pair relative to these end anchors and to assign a numeral between 1 and 20 to represent the appropriate position relative to the end values. The 34 *S*s rated the 25 adjective pairs of Replicate 2.

For the linemark response, *S*s were instructed to indicate each judgment of likableness by making a short vertical mark on a line so that the length between the margin and the mark would be proportional to the likableness. The 46 *S*s in this condition judged the adjective pair of Replicate 2.

The list for the matched pairs procedure was constructed from Anderson's (1968a) normative data. Each pair contained 2 adjectives of equal scale value. The 22 pairs covered the baseline range .35–5.60, in .25-category steps. The 50 *S*s were instructed to judge each of the adjective pairs of Replicate 2 by selecting the pair of adjectives from the list of 22 pairs, "most nearly equal in likeableness to the pair you are judging." The value of *S*'s response was taken to be the normative value of the adjective pair selected by *S*. The 22 pairs are listed in Birnbaum (1972b).

Supplementary scaling. These 22 adjective pairs and the 30 adjectives used in Experiment I were printed in random order. Each of 100 *S*s judged 3 aspects of each adjective or pair: (*a*) the *likableness* of a person who would be described by the adjective or pair, (*b*) the *activity* of such a person, and (*c*) the *range* of likableness of persons who would be described by the adjective(s). The *S*s judged all of the adjectives on one aspect before proceeding to the next task. Nine-point scales were used for all 3 scaling tasks, with 9 = like very very much, very very active, or very very wide range of possibilities.

Results and Discussion

Figure 3 shows the mean responses, averaged across *S*s, in each condition.

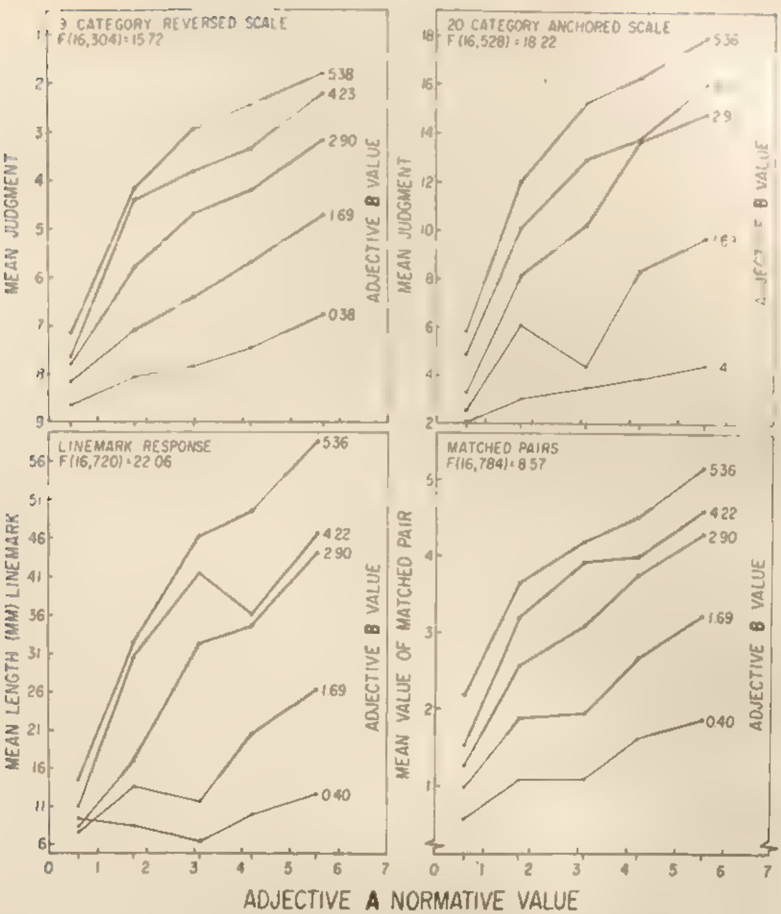


FIGURE 3. Mean judgments of likableness of pairs of adjectives: Experiment II. (Each panel shows results for a different response procedure.)

Additive and constant-weight averaging models again predict parallel curves for each panel. Instead, the interactions are all of the same general form (divergence) as in Figure 2, and the F values listed in each panel show that they are all highly significant.

The ordinate of the upper left-hand panel of Figure 3 has been relabeled for direct comparison with Figure 2. The data show the same interactions, and ratings of "dislikableness" appear to be a linear function (with negative slope) of ratings of "likableness." Since reversing the scale did not reverse the interaction, the deviation from additivity cannot be attributed to anything like preference for smaller numbers. The interaction for the anchored scale (upper right-hand panel) shows the same divergent form as those shown in the other panels and is highly significant. The percentage of variance associated with the interaction is actually larger for the linemarks condition than for any of the other conditions. The positively accelerated relationship between linemarks and category ratings may account for the larger interaction.

The matching procedure assumes that the normative values obtained by Anderson (1968a) are valid estimates of scale values. One check on this technique is to see whether the pairs having minimal within-set range are linearly related to Anderson's scale values. As can be seen by connecting these points (the diagonal) in the lower right-hand panel of Figure 3, this assumption is supported. Never-

theless, the interaction has the same divergent form as the others and is highly significant. The results for this nonnumerical response technique do not support the objection that perhaps *S*'s integration strategy is affected by the numerical response procedures typically used in this type of study. The error terms for this method were greater than for any of the other methods, perhaps because *S*s must judge the pairs of items defining the scale in addition to the usual stimulus pairs.

In summary, the main conditions of Experiment II show that variation of experimental procedures for responding did not eliminate the divergent interaction. Each procedure can be considered an alternative operational definition of the impression. By the principle of converging operations (Garner et al., 1956), the divergent interactions in Figure 3 support the interpretation that *impressions* of personality are nonadditive. In order to retain the additive model, it would be necessary to assume that the responses for all of these procedures are nonlinearly related to the impressions.

Supplementary scaling. One purpose of the supplementary scaling was to check a previous prediction (Birnbaum, 1972a) that items of lower value have narrower dispersions. Successive interval scale values and dispersions (Torgerson, 1958) were obtained for each adjective and adjective pair. Mean ratings were linearly related to Anderson's normative values and to Thurstone scale values derived from the same data. In agreement with the distributional interpretation, Thurstone scale values were positively correlated with dispersion values ($r = .65$). Since each of these dispersions was produced by the differences between the responses of different *S*s to the item, they are probably not the best estimates of the distribution of meaning for the adjectives. Only if it is assumed that different *S*s select a randomly sampled "person" from the distribution of "persons" who possess the trait would the Thurstone dispersion reflect this dispersion directly. Nevertheless, the correlation between value and dispersion seems consistent with this prediction.

Subjective estimates of the range of likableness seem a more direct measure of dispersion. These were highly correlated with mean ratings of likableness ($r = .91$). The plot of this relationship showed that although range is a nearly linear function of likableness for the dislikable traits, neutral and positive traits are all seen as implying similar wide ranges of likablenesses. This finding is consistent with the fact that the interactions in Figure 2 appear to be located in combinations that include dislikable traits. In a post hoc analysis of the 2 × 2 subdesign containing only positive adjectives for Experiment I, the interaction was nonsignificant ($F < 1$). The 2 × 2 subdesign containing the negative traits had a significant interaction, $F(1, 297) = 5.93$.

In summary, the supplementary scaling suggests that adjectives should be represented by distributions, with the lower valued items having smaller variance.

Supplementary test of activity. One interpretation of the crossover interaction observed in Experiment I, Replicate 3 (Figure 2) is that the activity component of the meaning of one adjective can multiply the evaluative component of the adjective with which it is paired. Hence, a SELF-CONFIDENT MALICIOUS person is more actively malicious than a SHY one. Similarly, an IMPULSIVE CRUEL person may be more likely to act cruel than an UNTIDY one (Experiment I, Replicate 2).

This hypothesis—that one adjective can behave like an adverb—was further investigated by having 130 additional *S*s rate the likableness of 36 hypothetical persons, each described by 2 adjectives produced from a 6 × 6 factorial design. The adjectives for the first factor varied in likableness (RUDE, IMMATURE, TROUBLED, DIRECT, INTELLECTUAL, and HONEST). The adjectives for the second factor were relatively neutral in likableness, but 3 were "active" (IMPULSIVE, AGGRESSIVE, and CHANGEABLE) and 3 were "passive" (QUIET, HESITANT, and SHY).

Consistent with the prediction, the likableness effect of an adjective was greater when paired with an active adjective. The interaction was large and significant,

F (25, 3225) = 14.49, and even showed reliable crossovers; for example, HESITANT & HONEST is less likable than AGGRESSIVE & HONEST although HESITANT & RUDE is more likable than AGGRESSIVE & RUDE.

Ratings of the activity of the 30 adjectives used in Experiment I were uncorrelated with ratings of likableness of the same adjectives but were correlated with the Thurstone dispersions (r = .43). The activity hypothesis is not sufficient to account for the overall divergent interactions in Figures 2 and 3, but it appears to explain some of the second-order effects (which would otherwise be called "peculiarities") for particular adjective combinations in Experiment I.

Experiment III: Subtractive Preference versus Additive Combination

Experiment III had the following objectives: (*a*) to evaluate a subtractive model for preference judgments reflecting the *difference* in likableness between 2 hypothetical persons, each described by one adjective; (*b*) assuming that the subtractive model were to fit, to compare scale values for the preference task with those obtained for the usual *combination* task, in which *S*s judge the likableness of a person possessing both traits; (*c*) to use the scale values obtained for the preference task as the basis for rescaling the data from the combination task; (*d*) to use the scale values for the preference task to evaluate Anderson's normative values for single items; and (*e*) to evaluate the effects of a change in instructions for the combination task, in which the adjectives are given equal importance and accuracy.

The subtractive model asserts that preference ratings can be represented as the algebraic differences between the values of the items:

$$\Psi_{ij}{}^{\mathrm{D}} = s_i - s_j, \qquad [2]$$

where $\Psi_{ij}{}^{\mathrm{D}}$ is the psychological difference in likableness (preference for Stimulus i over j), and s_i and s_j are the scale values of these 2 stimuli. If the subtractive model can be fit to the data, then the adjectives can be located as points o[illegible] a unidimensional likableness continuum [illegible]

Two principles of scale converge[illegible] be applied to this experime[illegible] sponse scale invariance: Wit[illegible] set of stimuli and the same r[illegible] cedure, the judgment function [illegible] to be independent of task. (*b*) St[illegible] scale invariance: With the sam[illegible] stimuli, the scale values of the stim[illegible] assumed to be independent of task. [illegible] ciple *a* implies that whatever transforma[illegible] tion is applied to fit the data to [illegible] ditive model for the combination [illegible] should also be applied to the data [illegible] the difference task. Principle *b* impli[illegible] that irrespective of whatever transformations of the responses fit the data for the 2 tasks to their respective models, the scale values derived from the subtractive model should be linearly related to the scale values for the same adjectives in combinations. Stimulus scale invariance can be considered a necessary (but not sufficient) condition for establishing meaningful scales and psychological laws. A failure of scale convergence can provide the leverage for rejecting either the subtractive or additive model, even though the J functions are unknown.

Method

Stimuli. The stimuli were the adjective pairs of Experiment I.

Subjects. The *S*s were 90 University of California, Los Angeles, undergraduates, 30 serving in each replicate.

Procedure. Each *S* performed both tasks, with half of the *S*s in each replicate performing the difference task first. There were no effects of task order.

In the difference task, *S*s were instructed to imagine 2 different people, each described by one of the adjectives of each pair, and then to judge the difference in likableness between the 2 persons. The 9-point scale had labels varying from 1 = like the person (described by the adjective) on the left very very much more than the person (described by the adjective) on the right, through 5 = like both persons equally, to 9 = like the person on the right very very much more than the person on the left.

The procedure for the combination task was in all respects identical to that of Experiment I, except that the instructions were modified to emphasize that the adjectives should be considered

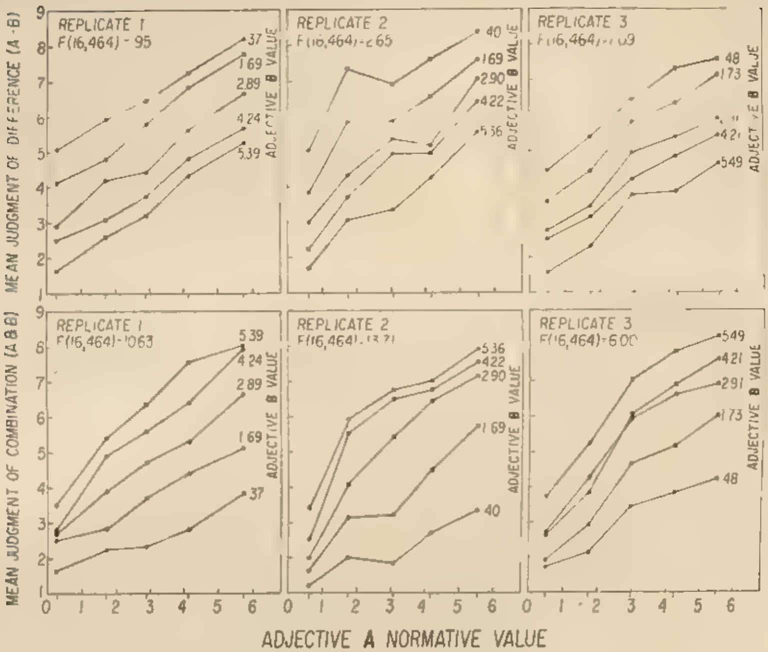


FIGURE 4. Mean ratings of difference in likableness between 2 hypothetical persons, each described by a single adjective (upper panels), and mean ratings of likableness of hypothetical persons, each described by the same 2 adjectives (lower panels): Experiment III.

of equal importance and accuracy, each word having been contributed by a different acquaintance "who knows the person well."

Results and Discussion

Difference task. The upper panels of Figure 4 show the mean judgments of the difference in likableness. Because ratings reflect the difference between Factors A and B, the slopes are positive, but the order of the curves is negative.

Since the subtractive model is a special case of an additive model, the parallelism of the curves supports its validity. Only the interaction for Replicate 2 is significant, and examination of the figure indicates that the interaction is minor and mainly due to one point. Since Anderson's (1968a) normative values are plotted on the abscissa, the near linearity of these functions supports the validity of the normative values as estimates of scale values. Similarly, the vertical separations between the curves are nearly proportional to differences in Anderson's scale values. The good fit of the subtractive model supports the validity of the category rating scale and indicates that the adjectives can be located on a unidimensional likableness continuum.

Combination task. The lower panels of Figure 4 show the mean judgments of the likableness of persons described by both adjectives. The data appear to replicate Experiment I closely, including the peculiarities for ratings of particular adjective combinations. The interactions for this experiment diverge in the same fashion as Experiment I and are highly significant for each replicate, as indicated by the F values printed in the panels of the figure. The F values are smaller than those for Experiment I because there were fewer Ss in this experiment. The change in instructions, emphasizing equal importance and accuracy, appears to have had no discernible effect on the interactions.

Scale convergence. Both the stimulus and response scale invariance assumptions would require that either the subtractive

model for differences or the constant-weight averaging model for combinations must be rejected. Response scale invariance implies that any transformation of the ordinate of the lower panels must be applied to the upper panels as well. Because the subtractive model fit the data directly, any nonlinear transformation would induce an interaction. Therefore, if the validity of the subtractive model and the principle of response scale invariance are assumed, the constant-weight averaging model must be rejected.

The fit of the subtractive model implies that the marginal means for these data constitute an interval scale for the adjectives (Anderson, 1970). However, when the combination data were rescaled to additivity using Kruskal's (1965) MONANOVA, scale values derived from this procedure were nonlinearly related to those for the same adjectives derived from the subtractive model. Therefore, the principle of stimulus scale invariance implies that this transformation would *not* be appropriate.

Further support for the validity of the rating scale is provided by the finding that ratings of homogeneous combinations (adjectives of similar normative value) are nearly linearly related (slightly sigmoidal) to subtractive model scale values. Hence, the assumption of stimulus scale invariance implies that the J function is roughly independent of task. Thus, 3 procedures agree—Anderson's (1968a) normative values, subtractive model scale values, and ratings of homogeneous combinations are all nearly linearly related.

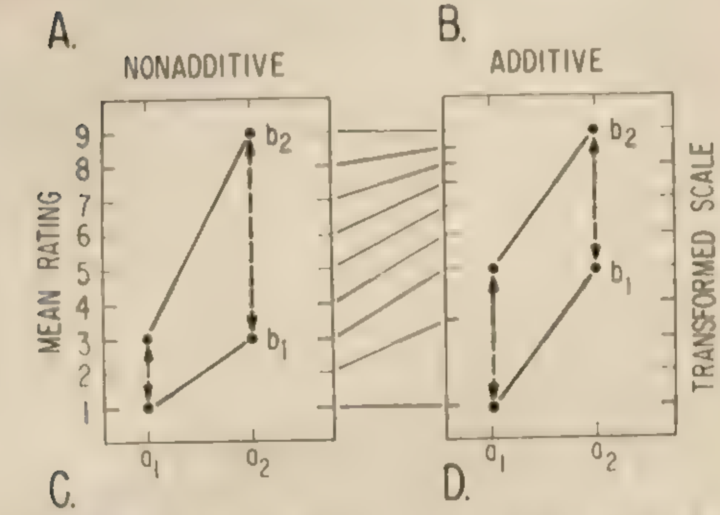
This analysis also indicates that a simple multiplicative model, $\Psi_{ij} = s_i s_j$, would be inappropriate for the combination task; it would incorrectly predict that judgments of adjectives of equal value are a positively accelerated quadratic function of the scale values for the difference model. However, a geometric averaging model (square root of the product of the scale values) would still be consistent with the subtractive model scale values.

The data of Experiment III indicate that with the same stimuli, *S*s, and general experimental procedure, ratings of [illegible]r-ence are consistent with a su[illegible] model, but ratings of combinatio[illegible] again inconsistent with the additive mo[illegible]ls. This suggests that the intera[illegible] due to trivial details of experim[illegible]cedure. If scale values for the adje[illegible]s are assumed to be independent of [illegible]k, the fact that the ratings of homoge[illegible]s combinations are a nearly linear fu[illegible]on of the scale values for the difference ra[illegible]s implies that it would be inappropria[illegible] rescale the ratings of combination[illegible] order to retain the additive mo[illegible] would be necessary either to reje[illegible] subtractive model (in spite of its goo[illegible] to the data in Figure 4), *or* assume t[illegible] *both* the scale values *and* the respo[illegible] function depend upon the task.

Experiment IV: Qualitative Nonadditivity

Experiment IV was designed to provi[illegible] a test of additivity that would requir[illegible] only the ordinal information in the data. The evidence against the additive and constant-weight averaging models obtaine[illegible] in the first 3 experiments is illustrated [illegible] Figure 5A, where the difference in rati[illegible]g between a_2b_2 and a_2b_1 is greater than t[illegible]e difference between a_1b_2 and a_1b_1. An a[illegible]ditive model requires that these differences be subjectively equal. If the J function is assumed to be linear, then differences in judgment are proportional to differences in the impressions; therefore, the divergence (previously shown in Figures 2, 3, and 4) would be contrary to additive models. However, if J were positively accelerated, impressions might be additive. The same ratings have been transformed to parallelism in Figure 5B. Although the difference in rating (Figure 5A) due to the change from b_2 to b_1 depends on whether a_1 or a_2 was in the same set, the difference in psychological value (Figure 5B) is assumed by additive models to be constant.

Experiment IV tested this possibility by asking *S*s to judge directly the differences in likableness between integrated impressions. The success of the subtractive



C.

	a_1b_1	a_1b_2	a_2b_1	a_2b_2
a_1b_1	5	6	6	9
a_1b_2	4	5	5	8
a_2b_1	4	5	5	8
a_2b_2	1	2	2	5

D.

	a_1b_1	a_1b_2	a_2b_1	a_2b_2
a_1b_1	5	7	7	9
a_1b_2	3	5	5	7
a_2b_1	3	5	5	7
a_2b_2	1	3	3	5

FIGURE 5. A. Schematic diagram of nonadditive data. B. Schematic [illegible] difference between the ordinate values for Panels A and B.]

model in Experiment III suggests that it can be used to infer the impressions of the combinations. It is only necessary to assume that ratings of differences in likableness are strictly monotonic to differences in psychological impressions. If the interaction is real, then the preference for a_2b_2 over a_2b_1 should exceed the preference for a_1b_2 over a_1b_1, as illustrated by the length of the dashed arrows in Figure 5A. If the interaction reflects only a nonlinearity in the J function, however, then the 2 preferences should be equal, as shown by the dashed arrows of Figure 5B.

The matrices in the lower panels of Figure 5 illustrate 2 patterns of data that might be obtained for ratings of the difference in likableness between 2 persons, each described by a pair of adjectives. Figure 5C shows the pattern that would be obtained if the integration of information were nonadditive. Each pair of circled entries is an example of a predicted rank-order comparison. The dashed ar-

TABLE 1

MEAN RATINGS OF THE DIFFERENCE IN LIKABLENESS BETWEEN PAIRS OF HYPOTH[illegible]

Person on the left	Person on the right			
	PHONY & MEAN	PHONY & TRUSTWORTHY	SINCERE & MEAN	[illegible]
Replicate 1				
PHONY & MEAN	4.91	6.34	6.31	7.[illegible]
PHONY & TRUSTWORTHY	3.83	5.07	4.94	[illegible]
SINCERE & MEAN	3.54	4.89	4.94	[illegible]
SINCERE & TRUSTWORTHY	1.66	2.60	2.51	[illegible]
	DECEITFUL & CRUEL	DECEITFUL & DEPENDABLE	HONEST & CRUEL	[illegible]
Replicate 2				
DECEITFUL & CRUEL	5.03	6.15	6.26	8.2[illegible]
DECEITFUL & DEPENDABLE	3.65	4.92	4.88	7.8[illegible]
HONEST & CRUEL	3.89	5.26	5.06	7.95
HONEST & DEPENDABLE	2.09	2.46	2.12	4.83
	MALICIOUS & OBNOXIOUS	MALICIOUS & UNDERSTANDING	LOYAL & OBNOXIOUS	LOYAL & UNDERSTANDING
Replicate 3				
MALICIOUS & OBNOXIOUS	5.03	6.35	6.37	[illegible]
MALICIOUS & UNDERSTANDING	3.77	5.00	[illegible]	[illegible]
LOYAL & OBNOXIOUS	3.58	4.85	5.06	[illegible]
LOYAL & UNDERSTANDING	1.98	2.55	2.40	[illegible]

rows represent the same comparison as the dashed arrows in Figure 5A. Since the ratings represent judgments of differences between impressions labeled by the column and the row entries, the direction of the order is reversed below the diagonal. Figure 5D represents the type of pattern that would be obtained if the adjectives combined additively. Each pair of circled entries would be equal. Both matrices in Figure 5 were generated by assuming the subtractive model for preference (Equation 2). However, the rank order of the matrix entries would remain the same if subjected to a strictly monotonic transformation. Thus, the data can be rescaled to fit the subtractive model without precluding the ordinal test for the additive integration models for impression formation (Birnbaum, 1972b).

For example, if the interaction shown in Figure 2 is real, then the judged difference in likableness between someone who is UNDERSTANDING & LOYAL and someone who is UNDERSTANDING & MALICIOUS should exceed the judged difference between someone who is OBNOXIOUS & LOYAL and someone who is OBNOXIOUS & MALICIOUS. But if adjectives [illegible] integrated according to an additive [illegible] constant-weight averaging model, then the 2 differences would be equal.

Method

The *S*s were instructed to judge the [illegible] in likableness between pairs of hypothetical [illegible], each of whom was described by 2 adjectives. Two adjectives were printed on the left side of the page and 2 on the right side of the page. The *S*s were instructed to form impressions of the personalities of the 2 persons before judging the difference in likableness. The 2 adjectives describing each person were described as being of equal accuracy and importance; *S*'s task was to imagine that they were contributed by different, but equally reliable, sources. Judgments were in terms of a 9-point scale, labeled as in the difference task of Experiment III.

Subjects. The *S*s were 195 University of California, Los Angeles, undergraduates, 65 in each of 3 stimulus replicates.

Stimuli. The stimulus for each judgment consisted of 4 adjectives, 2 printed on the left side of the page and 2 on the right. The adjective pairs describing the "person on the left" were produced from a 3 × 3, A × B, factorial design; the adjective pairs on the right were constructed from a 2 × 2, C × D, factorial design. Each person on the left was combined with each person on the right producing a 3 × 3 × 2 × 2, (A × B) × (C × D), factorial design.

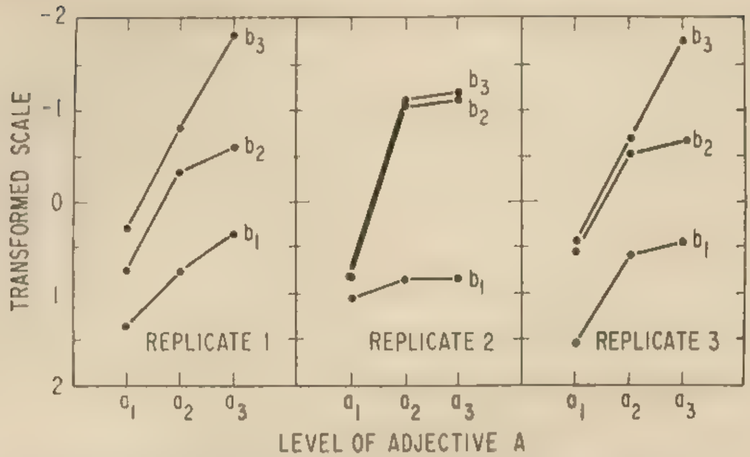


FIGURE 6. Transformed impression values for the person on the left derived from MONANOVA applied to difference ratings. Experiment IV (The 3 levels of subscripts refer to low, medium, and high likableness adjectives of Experiment I. For example, a_1 = PHONY, a_2 = BLUNT, and a_3 = SINCERE, for Replicate 1. To assess metric information in mean ratings, compare this figure with Figure 2.)

The basic idea of the stimulus construction can be understood by consideration of Figure 2. For the third replicate, the adjectives were a_1 = MALICIOUS, a_2 = CHANGEABLE, a_3 = LOYAL, b_1 = OBNOXIOUS, b_2 = SHY, b_3 = UNDERSTANDING, c_1 = MALICIOUS, c_2 = LOYAL, d_1 = OBNOXIOUS, and d_2 = UNDERSTANDING. With respect to Figure 5C, the last entry in the first row (9) would represent the rating of the difference in likableness between someone who is LOYAL & UNDERSTANDING and someone who is MALICIOUS & OBNOXIOUS. Nine additional cells in the design were constructed for each replicate by pairing each person on the left with an additional person on the right who was INFORMAL & LIGHTHEARTED. Therefore, the entire design represents a 9 × 5 factorial design with the factors representing the person on the left and the person on the right (see Birnbaum, 1972b).

Procedure. The 45 trials were printed in 1 of 3 different random orders in booklets with different page orderings for different *Ss*. The *Ss* read through the entire list before beginning to record their judgments.

Results and Discussion

The mean judgments of difference in likableness are presented in Table 1 for that part of the design illustrated in Figure 5. Comparison of the rank order of the ratings with the direction of the differences for the circled values shows that the pattern resembles that of Figure 5C representing an interaction: Differences in rating from Experiments I, II, and III predict the ratings of differences in the present experiment.

The crucial comparisons (circled entries in Figure 5D), which must be equal for the additive or constant-weight averaging model, are significantly in the direction predicted by the interactions in Figures 2–4 for all 3 replicates, Fs (1, 64) = 71.13, 115.88, and 89.54, respectively. The individual ratings by single *Ss* show the same pattern of rank orders for these crucial comparisons as the mean judgments in the table. Of the 195 *Ss*, 81 were in perfect ordinal agreement with the interaction obtained in the first 3 experiments for all 4 comparisons. Only 12 *Ss* showed a greater number of rankings in the direction opposite from that predicted by the interaction, compared with 170 *Ss* whose rankings were in the direction predicted by the interactions obtained in Experiment I. These results are clearly inconsistent with the additive model which would predict an equal split of these rankings between the 2 directions.

The larger, 9 × 5 design was used to test the subtractive model for preferences between combinations. The subtractive

model also predicts parallelism, and the data appeared roughly parallel when plotted. However, statistically significant interactions were obtained for each stimulus replicate, Fs $(32, 2048) = 3.72$, 4.98, and 5.10 for Replicates 1, 2, and 3, respectively. Application of MONANOVA to these ratings of preference reduces the interaction to less than 1% of the total variance, so it is apparent that the preference ratings are at least ordinally consistent with the subtractive model.

Following the application of MONANOVA to the 9×5 subtractive model design, the derived impression values for the combinations were inconsistent with the additive and constant-weight averaging models. Figure 6 plots the impression values derived from the MONANOVA transformation for the 9 persons on the left. Since these were produced from a 3×3 factorial design of adjectives, the nonparallelism of these scale-free values refutes the additive models of impression formation. Similar results were obtained for the 2×2 design for persons on the right.

In summary, the minimal assumption that the preference ratings are monotonically related to subjective differences leads to the conclusion that the additive and constant-weight averaging models must be rejected. Comparison of scale-free values in Figure 6 with the scale-dependent ratings in Figures 2, 3, and 4 shows that ratings contain metric information that correctly describes the divergent interaction.

Krantz et al. (1971, pp. 445–447) have argued that if an interaction can be removed by monotonic transformation, then it should be attributed to the rating scale and should not be given a psychological interpretation. However, Experiment IV shows that this procedure could lead to erroneous conclusions. Although the interaction obtained in Experiment I could be removed by monotonic transformation, ratings of differences in impression are predictable from the differences in ratings. This indicates that the interactions are of psychological significance. Rescaling those data to parallelism would be inappropriate, as demonstrated by the qualitative contradictions of Experiment IV.

However, the ratings also appear to contain some nonlinearity that reduces the apparent magnitude of the interaction for nonnegative items. It would be improper reasoning to conclude that nonnegative items combine additively based on the lack of interaction in the ratings. The scale-free values in Figure 6 appear to show a continued divergence. Nonlinearity in the response scale may actually make nonadditive data appear additive. The procedures of Experiment IV avoid the rescaling problems and provide a truly scale-free test of the additive models.

General Discussion

This research demonstrates that impressions of likableness cannot be represented as simple sums or constant-weight averages of the values of the adjectives. Instead, one bad trait results in an unfavorable overall impression with the other trait having less influence. This effect reflects a real psychological interaction and cannot be attributed to the response scale.

Some of the early research on impression formation concluded that the constant-weight averaging model could give a reasonable fit to ratings of adjective combinations. The parallelism prediction appeared to be roughly satisfied, and this finding was interpreted to validate both the model and the rating scale. Systematic deviations did occur in this research, but were often attributed to experimental difficulties or nonlinearity of the response scale. However, it is now clear that the constant-weight averaging model is not an appropriate general description, but may give a reasonable approximation when the stimulus range is restricted and the experimental design lacks power.[3]

[3] An alternative view might maintain that the constant-weight model is descriptive of impression formation, but only under limited experimental circumstances. The E would be advised to carefully select adjectives, instructions, response procedures, and other conditions to minimize the nonparallelism. When the data appear parallel, this view contends that E would have the right to conclude that the parallelism jointly supports the constant-weight model and the response scale. However, this approach seems unsatisfactory for several reasons. First, the present research shows that different response procedures, "equal impor-

Theoretical Implications

It is useful to consider a set of conditions that yield the parallelism prediction to consider what nonparallelism might mean. The following conditions underlie the parallelism prediction of the averaging model: (*a*) the integration function is an averaging process; (*b*) the adjectives within each factor have equal absolute weight; (*c*) the adjectives do not change value nonlinearly in combination, nor do the weights depend upon the particular stimulus configuration; and (*d*) the J function is linear. Based on a single experiment such as Experiment I, nonparallelism could be interpreted as evidence to disprove at least one of the premises. Without further constraints, it would be impossible to specify whether nonparallelism was due to a nonadditive I process (Conditions *a–c*) or a nonlinear J function (Condition *d*). These experiments provide the leverage to indicate that the nonparallelism is *not* attributable to the J function. There are several remaining possibilities: The integration function may not be an averaging process, the stimulus parameters may depend upon the configuration, or the weights may not be equal for the adjectives.

The differential-weight averaging model (Anderson, 1971, 1972, 1974) can account for the interactions by allowing the weight of an item to depend upon its scale value. Differential weighting requires many more parameters and seems unnecessarily complicated to fit the simple divergence interactions. The model can account for a wide variety of interactions, making it difficult to disprove. However, with large enough designs, there are degrees of freedom left over to permit tests (e.g., Anderson, 1972; Birnbaum, 1973). Configural weighting (Birnbaum, 1972b), in which the weight of an item depends in part on its rank within the set, requires fewer parameters and gives a slightly better fit to the data.[4]

The configural-weight model predicts steady divergence for Figures 2, 3, 4, and 6. The differential-weight model is more flexible and could account for reconvergence, among other patterns. It is thus of theoretical interest to ask if the curves show any evidence of reconvergence. The scale-dependent ratings of Figure 2 appear nearly parallel for nonnegative traits. However, the scale-free values in Figure 6 show steady divergence for Replicates 1 and 3. The ratings may contain a small scale-end effect that reduces the apparent interaction. The bulge in Figure 2, Replicate 2 is apparently due to IMPULSIVE multiplying the effect of the other adjective, rather than having less weight on its own. Since there is no evidence for more than a simple divergence, the existing data cannot test between differential and configural-weight models. Methods for distinguishing these models have been suggested by Birnbaum (1973).

Methodological Implications

Previous conceptualizations of the stimulus concatenation problem have not separated the

tance and accuracy" instructions, and different selections of adjectives all yield similar divergent interactions. If the model is to be deemed correct, then it must apply under very special conditions indeed. Second, since it is possible to manipulate the form of J, it follows that selection of experimental procedures could lead to an "experimental rescaling" of the data. Thus, it should be possible to select end anchors and filler stimuli to reduce the interaction in the analysis of variance. Therefore, it would be inappropriate to select experimental procedures that yield parallel data and then conclude that the parallelism is joint support for model and response scale. That parallelism could result from a combination of nonadditive I and nonlinear J is not merely an untestable philosophical possibility but represents a plausible hypothesis that can be tested by the procedures of Experiment IV.

[4] An averaging mechanism can be analogized to a balance plank and fulcrum. In the differential-weight model, each adjective corresponds to a weight (w) placed at a certain location (s) on the plank. The integrated impression (Ψ) is the location where the fulcrum must be placed so that the plank will balance. This location is the weighted average, $\Psi = \sum w_i s_i / \sum w_i$. The configural-weight averaging model assumes that the weight of a stimulus depends upon its rank within the set to be judged. For 2 stimuli, the simple range model (Birnbaum et al., 1971), $\Psi_{ij} = .5(s_i + s_j) + \omega|s_i - s_j|$, can be rewritten as a configural-weight model, $\Psi_{ij} = (.5 + \omega)s_i + (.5 - \omega)s_j$, when $s_i > s_j$. This model can also be represented by a balance plank mechanism; however, the configural-weight model does not require each location on the plank to have its own weight. For 2 stimuli in the set, there would be only 2 weights, one for the lower valued item, and one for the higher valued item. The same weight can be placed at any location on the plank, if it holds the same rank. This model becomes a *minimum* model when $\omega = -.5$, a *constant-weight* model when $\omega = 0$, and a *maximum* model when $\omega = .5$, and can describe a family of simple convergent or divergent interactions.

J and I functions of Figure 1. This article offers a conceptualization that separates these problems. The propriety of rescaling the data has been uncertain in previous work. Functional and conjoint measurement (Anderson, 1970; Krantz et al., 1971) both allow for monotone transformation, but differ in their outlook about when transformation is appropriate.

Krantz et al. (1971, pp. 445–447) reanalyzed the data of Sidowski and Anderson (1967) and concluded that since the interaction could be eliminated by a monotonic transformation of the mean ratings, the original authors were incorrect in attributing psychological significance to their findings. An investigator following this rescaling procedure with the data of Experiment I would have erroneously concluded that the additive model was an appropriate description of impression formation. The present research, by demonstrating that data transformation can lead to erroneous conclusions, provides a warning against this practice.

Anderson[5] originally attempted to rescale the data for ratings of the severity of disturbed behaviors and to attribute the nonadditivity to the rating scale, but he has recently reinterpreted the same data as evidence for the configurality of clinical judgment by fitting these data to an averaging model with differential weights (Anderson, 1972). The latter interpretation assumed the linearity of the J function.

The present research supports this attitude toward rating scales, but it also provides scale-free constraints that make the interpretations of Krantz et al. (1971) and Anderson (1972; see, also, Footnote 5) matters for experimentation rather than assumption.

In Experiment IV it would be possible to distinguish multiplicative from additive models on the basis of ordinal information (see Figure 5). The simple view of conjoint scaling would not allow this distinction, since a monotonic (logarithmic) transformation of a positive product yields a sum. By assuming that ratings of differences are monotonically related to subjective differences, the I function can be separated from the J function. Experiment IV allows rescaling, but rescaling does not preclude the possibility of testing the I function.

Although the rating scale contains metric information, it does not seem appropriate to assume that J is linear in every experiment. The J function can be predictably nonlinear. It depends upon contextual effects that are explainable (Birnbaum et al., 1971; Parducci & Perrett, 1971). A nonlinear J function could result in 2 possible errors: (a) nonadditive data when the underlying impressions are additive, or (b) additive data when the impressions are nonadditive. Thus, the decision to rescale or not to rescale cannot be convincingly settled without further constraints, such as those applied in Experiments III and IV.

Conclusions

Impressions of likableness cannot be represented as simple sums or averages of single values of the adjectives. They appear to be a predictable, but nonadditive function of the component values. The data show consistent, regular deviations from additivity that are similar for different selections of adjectives. When one adjective is dislikable, the person is rated as dislikable, and variation of the other trait has less effect.

Differential or configural weighting of the more dislikable traits can account for the interactions. Representation of the stimuli by distributions could explain why the adjectives are integrated by a nonadditive function: It is less likely for a person possessing a dislikable trait to be likable than for a person with a likable trait to be dislikable.

Finally, this research illustrates that rating scales contain metric information that should not be uncritically scaled away. The procedures employed in Experiments III and IV provide constraints that allow a model to be tested without having to assume the validity of the rating scale.

[5] N. H. Anderson. Application of an additive model to impression formation. Paper presented at the third annual meeting of the Psychonomic Society, St. Louis, August 1962.

REFERENCES

ANDERSON, N. H. Application of an additive model to impression formation. *Science*, 1962, **138**, 817–818.

ANDERSON, N. H. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 1968, **9**, 272–279. (a)

ANDERSON, N. H. A simple model for information integration. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968. (b)

ANDERSON, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170.

ANDERSON, N. H. Integration theory and attitude change. *Psychological Review*, 1971, **78**, 171–206.

ANDERSON, N. H. Looking for configurality in clinical judgment. *Psychological Bulletin*, 1972, **78**, 93–102.

ANDERSON, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 7. New York: Academic Press, 1974.

ANDERSON, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*. Vol. 2. New York: Academic Press, in press.

ASCH, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 1946, **41**, 258–290

BIRNBAUM, M. H. Morality judgments: Tests of an averaging model. *Journal of Experimental Psychology*, 1972, **93**, 35–42. (a)

BIRNBAUM, M. H. The nonadditivity of impressions. Unpublished doctoral dissertation, University of California, Los Angeles, 1972. (b)

BIRNBAUM, M. H. Morality judgment: Tests of an averaging model with differential weights. *Journal of Experimental Psychology*, 1973, **99**, 395–399.

BIRNBAUM, M. H., PARDUCCI, A., & GIFFORD, R. K. Contextual effects in information integration. *Journal of Experimental Psychology*, 1971, **88**, 158–170.

BIRNBAUM, M. H., & VEIT, C. T. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, 1974, **15**, 1–9.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychological Review*, 1956, **63**, 149–159.

KANOUSE, D. E., & HANSON, L. R., JR. *Negativity in evaluations*. Morristown, N.J.: General Learning Press, 1972.

KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. *Foundations of measurement*. New York: Academic Press, 1971.

KRANTZ, D. H., & TVERSKY, A. Conjoint measurement analysis of composition rules in psychology. *Psychological Review*, 1971, **78**, 151–169.

KRUSKAL, J. B. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society* (Series B), 1965, **27**, 251–263.

KRUSKAL, J. B., & CARMONE, F. J. MONANOVA: A FORTRAN-IV program for monotone analysis of variance. *Behavioral Science*, 1969, **14**, 165–166.

LAMPEL, A. K., & ANDERSON, N. H. Combining visual and verbal information in an impression-formation task. *Journal of Personality and Social Psychology*, 1968, **9**, 1–6.

PARDUCCI, A., & PERRETT, L. F. Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology Monograph*, 1971, **89**, 427–452.

SIDOWSKI, J. B., & ANDERSON, N. H. Judgments of city-occupation combinations. *Psychonomic Science*, 1967, **7**, 279–280.

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

(Received November 16, 1972)

Manuscripts Accepted for Publication in the

Journal of Experimental Psychology

Body Temperature and Temporal Acuity. James F. O'Hanlon (Human Factor- [illegible] Barbara Research Park, Goleta, California 93017), James J McGrath, and [illegible]

Processing Distinctive Features in the Differentiation of Letter-Like Symbols. [illegible] Menlo Avenue, Glenside, Pennsylvania 19038) and Francine Simmons.

Effects of Positive and Negative Force-Contingent Reinforcement on the Frustration [illegible] Gail S. Ditkoff (Biddle Hall, University of Pittsburgh at Johnstown, Johnstown [illegible]) and Ronald Ley.

Imagery and Cued Recall: Concreteness or Context? Ronald C. Peterson (Clinical [illegible], Medical Research Division, Headquarters, Edgewood Arsenal, Aberdeen Proving Gr[illegible] 21010).

Regression Effect and Individual Power Functions over Sessions. William E. Dawson [illegible] of Psychology, University of Notre Dame, Notre Dame, Indiana 46556) and Robert G. [illegible]

The Relationship of Interval Frequency Count to Ratings of Melodic Intervals. Thomas [illegible] (10514 National Boulevard, Apartment 323, Los Angeles, California 90034).

Maintenance of Interference in Short-Term Memory. Judith Goggin (Department of Psycho[illegible] versity of Texas at El Paso, El Paso, Texas 79968) and Donald A. Riley.

The Linguistic Interdependence of Bilinguals. Robert K. Young (Department of Psychology, M[illegible] Hall 211, University of Texas at Austin, Austin, Texas 78712) and Mike Lopez.

Good Continuation Revisited. Lubomir S. Prytulak (Department of Psychology, University of Western Ontario, London 72, Ontario, Canada).

Haptic Equivalence Matching of Curvature by Blind and Sighted Humans. Philip W[illegible] (Division for Disorders of Development and Learning, Child Development Institute, Univ[illegible] Carolina, Chapel Hill, North Carolina 27514) and Teresa T. Whitson.

The Combined Effects of Stimulus and Response Conditions on the Delay in Identify[illegible] of Words. Raymond L. Majeres (Department of Psychology, Western Illinois [illegible], Illinois 61455).

Evidence of a Primary Frustration Effect following Quality Reduction in the Double R[illegible] Cross (Department of Psychology, Colorado State University, Fort Collins, Colorado 80521) and William N. Boyer.

Stimulus-Reinforcer Predictiveness and Selective Discrimination Learning in Pige[illegible] man (Department of Psychology, University of Iowa, Iowa City, Iowa 52240).

Memory Search of Categorized Lists: A Consideration of Alternative Self-Terminating [illegible] Mary J. Naus (Department of Psychology, Wellesley College, Wellesley, Massachusett[illegible]

Attenuation of Blocking with Shifts in Reward: The Involvement of Schedule-Generated Con[illegible] Allen R. Wagner (Department of Psychology, Yale University, 333 Cedar Street, New H[illegible] necticut 06510) and James H. Neely.

Partial Reward Either following or Preceding Consistent Reward: A Case of Reinforcement Lev[illegible] J. Capaldi (Department of Psychological Sciences, Purdue University, West Lafayette, Indian[illegible]7).

Effects of Category Attention, Relative Frequency of Relevant Values, and Practice on Attribute Ide[illegible]cation Performance. Peder J. Johnson (Department of Psychology, University of New Mexico, Albuquerque, New Mexico 87131) and Thomas C. Toppino.

Intralist Contrast Effects in Cued Recall. Donald Robbins (Department of Psychology, Emory University, Atlanta, Georgia 30322), James F. Bray, and James R. Irvin.

Proactive Interference and Directed Forgetting in Short-Term Motor Memory. Leslie Burwitz (Department of Education and Recreation, Acadia University, Wolfville, Nova Scotia, Canada).

Parallel vs. Sequential Processing of Pictures and Words. Joan G. Snodgrass (Department of Psychology, 4 Washington Place, Room 857, New York University, New York, New York 10003) and George Antone.

Restriction on the Ranschburg Effect. John C. Jahnke (Department of Psychology, Miami University, Oxford, Ohio 45056).

Parallel Processing in a Word-Nonword Classification Task. Neil Novik (Department of Psychology, University of Connecticut, Storrs, Connecticut 06268).

Effects of Familiarity and Sequence Length on Analog Matches in the Simultaneous Matching Task. Gail A. Bruder (Department of Psychology, State University of New York at Buffalo, 4230 Ridge Lea Road, Buffalo, New York 14226) and Wayne Silverman.

The Effects of Imagery Instructions, Imagery Ratings, and the Number of Dictionary Meanings upon Recognition and Recall. M. J. Peterson (Department of Psychology, Indiana University, Psychology Building, Bloomington, Indiana 47401) and S. H. McGee.

Syllable-Dependent Pronunciation Latencies in Number Naming: A Replication. Stuart T. Klapp (Department of Psychology, California State University, Hayward, California 94542).

Recall Accuracy of Eidetikers. Charles J. Furst (Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755), Kenneth Fuld, and Michael Pancoe.

Presentation Rate and Intralist Repetition Effects in Immediate Probed Recall. V. David Burns (c/o Nancy S. Anderson, Department of Psychology, University of Maryland, College Park, Maryland 20740).

Estimation of Word Frequency in Continuous and Discrete Tasks. Ian Begg (Department of Psychology, McMaster University, Hamilton, Ontario, Canada).

CONTENTS—(*Continued*)

Journal of Experimental Psychology

CONTENTS

ARTICLES

(*Continued on inside back cover*)

American Psychological Association

Vol. 102, No. 4

April 1974

Beginning in 1975 the *Journal of Experimental Psychology* will appear as four independently edited and distributed sections:

The **Journal of Experimental Psychology: Human Learning and Memory** (Lyle E. Bourne, Jr., Editor) will publish experimental studies on fundamental acquisition, retention, and transfer processes in human behavior.

The **Journal of Experimental Psychology: Human Perception and Performance** (Michael I. Posner, Editor) will publish experimental studies designed to foster understanding of information processing operations and their relation to experience and performance.

The **Journal of Experimental Psychology: Animal Behavior Processes** (Allan R. Wagner, Editor) will publish experimental studies concerning the basic mechanisms of perception, learning, motivation, and performance, especially as revealed in the behavior of infrahuman animals. The studies should make significant contributions to general behavior theory.

The **Journal of Experimental Psychology: General** (Gregory A. Kimble, Editor) will publish experimental articles in any of the three areas represented by the above sections, when the articles involve a longer, more integrative report leading to an advance in knowledge that is judged to be of interest to the entire community of experimental psychologists. The section will include but not be limited to articles such as those that have appeared as *Journal of Experimental Psychology* monographs and as chapters in contemporary books of "advances." Republication of a limited number of data may be permitted if necessary to make the article complete and definitive.

In some instances it may be difficult for the author to decide whether one of the above listed sections or the *Journal of Comparative and Physiological Psychology* is most appropriate for a given manuscript. The author should weigh the essential implications of the research rather than the specific techniques or subject population employed. Is the central message concerned with the issues and concepts of behavior theory (*JExP*), or with the biological substrates of behavior (*Journal of Comparative and Physiological Psychology*)? Is the study primarily addressed to theory of human behavior (*JExP: Human Learning and Memory* or *JExP: Human Perception and Performance*) or to theories equally or more relevant to infrahuman subjects (*JExP: Animal Behavior Processes*)? Is the major emphasis on the acquisition, retention, or transfer of new information, concepts, or problem-solving strategy (*JExP: Human Learning and Memory*) or on the organization or control of structures or mental operations, where the acquisition of new skills is only incidental (*JExP: Human Perception and Performance*)?

Manuscripts should be sent directly to the Editor of the relevant section at the address listed below. In cases where an Editor believes a manuscript is clearly more appropriate for an alternative journal of the American Psychological Association, he may redirect the manuscript with the approval of the author. A final action taken by one of the Editors will be final for all sections of *JExP*.

Manuscripts should be submitted in triplicate and according to the instructions contained in the *Publication Manual* of the American Psychological Association.

Authors should note that limitations on total pages make it necessary to restrict publication to those papers that in the opinion of the Editors make a notable contribution to the literature. Preference can be expected for sets of interrelated experiments that are definitive or provocative.

Addresses for the 1975 Editors-Elect are as follows: *JExP: Human Learning and Memory*, Lyle E. Bourne, Jr., Department of Psychology, University of Colorado, Boulder, Colorado 80302; *JExP: Human Perception and Performance*, Michael I. Posner, before June 1, 1974, same as Allan R. Wagner, after June 1, Department of Psychology, University of Oregon, Eugene, Oregon 97403; *JExP: Animal Behavior Processes*, Allan R. Wagner, Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510; *JExP: General*, Gregory A. Kimble, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1974: $60 domestic, $61 foreign, $6 single issue. APA member rate: $25. Write to Subscription Section, APA.

Reprints: Authors may order reprints of their articles from the printer when they receive galley proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., 175 Fifth Avenue, New York, New York 10010; University Microfilms, Ann Arbor, Michigan 48106, or Princeton Microfilm, Princeton, New Jersey 08540.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. Undelivered copies resulting from address changes will not be replaced; second-class mail is not forwarded unless subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Published monthly in two volumes, six issues per volume, by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Washington, D.C., and additional mailing offices.

Journal of Experimental Psychology

Vol. 102, No. 4 — April 1974

CARDIAC ORIENTING DURING "GOOD" AND "POOR" DIFFERENTIAL EYELID CONDITIONING[1]

LOIS E. PUTNAM, LEONARD E. ROSS, and FRANCES K. GRAHAM[2]

University of Wisconsin—Madison

The relation between orienting responses (ORs) and acquisition of a differentially conditioned eye blink response was studied in a group of 10 *Ss* showing good differential conditioning and a group of 10 *Ss* who failed to condition differentially. With Good conditioners, the predictions of OR theory were confirmed. Orienting, measured by heart rate (HR) deceleration, increased during acquisition and habituated as conditioning stabilized. With Poor conditioners, an OR-interference effect was found which is consistent with OR theory but was unexpected on the basis of most previous work. Another unexpected result was the occurrence of HR accelerations which were larger to a signaled than to an unsignaled unconditioned stimulus (UCS). It was suggested that both findings could be explained by the effects of orienting on processes active during short as opposed to long CS-UCS intervals and on skeletal conditioned responses (CRs) as opposed to autonomic CRs that are components of the orienting system.

A major reason for interest in the orienting response (OR) is its presumed relation to learning processes. Pavlov initially believed that the pattern of overt physical reactions which he termed the "investigatory" or "orienting" reflex had little importance for learning, but he later noted that conditioning proceeds best in the presence of an OR if it is neither too strong nor too weak. He also noted that the OR tends to disappear as conditioning is established (Lynn, 1966; Razran, 1961). The more fully articulated theory of orienting developed by Sokolov (1963) hypothesizes two generalized systems of responses with opposed functions, an OR which enhances stimulus reception and a defensive response system (DR) which attenuates the effects of stimulation. The relation of these systems to changes occurring in learning was described for two paradigms—classical conditioning of the vasomotor response to shock and the conditioning of an instructed motor response. In both cases, a vasomotor OR appeared on initial pairings of conditioned stimulus (CS) and unconditioned stimulus (UCS), intensified on subsequent pairings, persisted until the conditioned response (CR) stabilized, and disappeared on later trials. After vasomotor conditioning to shock was accomplished, both CS

[1] This study was supported by Research Grants [illegible] and HD01490, by Research Scientist Award K3-MH21762, and by Training Grant HD00117 from the U.S. Public Health Service. Computer services were provided through Grant FR00249 to the Laboratory Computer Facility and through a National Science Foundation grant administered by the University of Wisconsin Research Committee. Portions of this paper were presented at meetings of the Society for Psychophysiological Research, St. Louis, Missouri, October 1971.

[2] Requests for reprints should be sent to Leonard E. Ross or Frances K. Graham, Department of Psychology, University of Wisconsin—Madison, Madison, Wisconsin 53706.

and UCS elicited a stable vasomotor DR. Since the instructed motor response situation did not employ an aversive UCS, a DR was not elicited and, after conditioning was stabilized, vasomotor responses disappeared entirely (Sokolov, 1963, pp. 172 173, 185; Vinogradova, 1965).

Findings obtained from these two paradigms constitute the primary evidence that an OR is a prerequisite for normal conditioning and that the OR changes during the course of conditioning. However, both paradigms have limitations. Instructions are a means of manipulating attention and may thus elicit a different OR pattern than would occur with the traditional conditioning situation. With an autonomic conditioning paradigm, the measures of OR and CR may not be independent. As Morgenson and Martin (1968) have pointed out, when the same response measure indexes both the OR and CR, relationships between them may be due more to general responsiveness within systems than to any role of the OR in conditioning. Although the occurrence of changes in the direction of cephalic vasomotor responses, depending on whether an OR or DR is elicited, may exempt the Soviet studies from this criticism, the existence of OR DR differentiation by vasomotor responses has not been unequivocally replicated by Western investigators (Graham, 1973). In any case, the conditioning studies of Western investigators who have been concerned with OR changes have mainly employed the galvanic skin response (GSR) or the peripheral vasomotor response, neither of which distinguishes between ORs and DRs. These studies are clearly vulnerable to the objection of possible OR and CR confounding unless latency criteria are used to separate the responses and, even then, to the objection that any OR-learning association is due to the reactivity of the particular response system. The objection is especially pertinent to studies concluding that ORs, [illegible]dexed by the GSR on initial trials, predict fac[illegible]ilitated GSR conditioning (e.g., Maltzman & Mandell, 1968).

[illegible]espite the obvious advantages of study- [illegible] and CR in separate response systems and of using a [illegible] of de- pendent on verbal instruct[illegible] [illegible]re ap- pear to be no studies of hu[illegible] [illegible] which a classically conditioned [illegible] motor response and an OR DR [illegible]tiating autonomic response have [illegible] [illegible]igated concurrently. This parad[illegible] [illegible]sed in the present study. The eye[illegible] [illegible]sponse to air puff was the to-be-co[illegible] [illegible]ed re- sponse, and orienting was [illegible] [illegible]ed by heart rate (HR) deceleration. [illegible]t rate deceleration appears to meet [illegible] [illegible]riteria of an OR and HR acceleration [illegible] [illegible]riteria of a DR (Graham, 1973; Jacks[illegible] 1972). Since acquisition of a differenti[illegible] [illegible]CR is less affected than a simple CR [illegible] gen- eralized arousal effects (Prokasy, 1[illegible]), the eye blink response was differentia[illegible] con- ditioned. While OR theory predicts rela- tively unambiguous OR changes in response to the positive CS (CS+) as conditioning progresses, it is less clear what changes in the OR should be expected wi[illegible] [illegible] nega- tive CS (CS−). If differen[illegible] [illegible]did con- ditioning is to occur, both [illegible] [illegible]ditioned stimuli should presumably [illegible] [illegible]ienting, and changes in orienting mi[illegible] [illegible] expected to follow a similar course, p[illegible]bly with less pronounced orienting an[illegible] [illegible]ore rapid extinction to the nonrein[illegible]d CS−. Sokolov (1960, p. 221) gives an [illegible]ample of more rapid extinction of ori[illegible]ng to the negative stimulus, but the sit[illegible]ion is not directly comparable to the pres[illegible] one since differential conditioning was preceded by simple conditioning.

In addition to investigating HR changes during differential eyelid conditioning, the present study also compared these changes with HR responses to the conditioned and unconditioned stimuli when they were unpaired. Finally, the relation of orienting to successful conditioning was studied by selecting two groups of *S*s who differed markedly in how well they achieved differential eyelid conditioning. If the OR plays a role in facilitating acquisition, "good" conditioners should show the expected pattern in response to CS+, i.e., initially increasing HR decelerations which would habituate as differential eyelid conditioning reached asymptote. "Poor" conditioners,

on the other hand, might show either unusually large and persistent ORs or smaller, less stable ORs. The latter finding would be in line with Maltzman's conclusion that low-OR *S*s are poorer conditioners (e.g., Maltzman & Mandell, 1968).

Method

Subjects. Fifty undergraduates, receiving experimental "points" for their participation, served as *S*s. Five additional *S*s were tested and discarded in order to complete groups of Good and Poor differential conditioners. Of these five *S*s, three failed to meet a criterion of producing at least five CRs, and two were tested to complete the Poor conditioning group, but showed good differential conditioning. Two other *S*s were also discarded, one due to equipment failure and one for a heart anomaly.

Apparatus. The subject-testing room contained an upright ophthalmologist's chair which faced a screen displaying the fixation point—a black, circular, cardboard disc 8.5 cm. in diameter. A rectangular patch of light (85 × 50 cm.), projected on the screen, provided the only illumination.

Recording and stimulus generating equipment were located in a separate room having intercom communication with the *S* room. Two Hewlett-Packard ABR audio oscillators (Model 200 CDR) produced pure tones of 800 and 2,100 Hz. whose rise and decay times ($\approx 10\ \mu s.$) were controlled by a Grason-Stadler Model 829C electronic switch. A Grason-Stadler 901B noise generator provided continuous background white noise which, with tones, was presented binaurally to *S* through Sharpe HA 10A earphones. Intensity of tones and noise was determined from matches of four judges who adjusted the earphone-presented noise to the loudness of free-field white noise, measuring 65 db. re .0002 microbars on the C scale of a General Radio 1551C sound-level meter, and adjusted the earphone-presented tones-over-noise to the loudness of a 1,000-Hz. free-field tone 85 db. above the judges' mean threshold. Stimulus durations, interstimulus intervals, and intertrial intervals (ITIs) were controlled by Tektronix timing equipment and an Ohrtronics tape reader. Tones were automatically triggered on the positive slope of the first electrocardiographic R-wave occurring after the ITI had elapsed.

Movement of the *S*'s right eyelid was recorded using a standard potentiometer–polygraph apparatus (Wilcox & Ross, 1969) and heart rate, detected by a lead II placement of Beckman biopotential electrodes, was tape recorded and computer read.

Design and procedures. Five treatment groups of five males and five females each, to which *S*s were assigned randomly until full, included two differential eyelid conditioning groups (Good and Poor conditioners) and three control groups (Unpaired, Two-Tone, and One-Tone). Subjects retained in

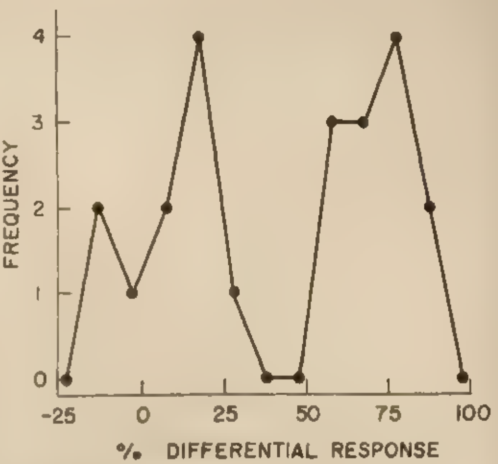


FIGURE 1. Distribution of differential eyelid conditioning performance.

the Good group showed 55% to 90% differential conditioning on the last half of the conditioning trials (percentage CRs on the last 20 CS+ trials less percentage CRs on the last 20 CS− trials), while *S*s in the Poor group ranged between −15% to 25%. The distribution of percentage difference scores for all *S*s tested, including the two discarded Good conditioners, appeared to be bimodal, as illustrated in Figure 1.

Conditioning *S*s received 40 CS+ and 40 CS− trials with an ITI of 18 to 30 sec. from CS onset to CS onset. A Gellerman presentation order, repeated every 20 trials, provided that each block of 10 trials contained equal numbers of CS+ and CS− trials and that neither stimulus occurred more than three times in succession. The UCS was a 100-msec. puff of .75 psi compressed air delivered to *S*'s right cornea and coterminous with the 900-msec. tone CS+. Six of the Good conditioners received a 2,100-Hz. CS+ and an 800-Hz. CS− and four received the reverse pattern. In the Poor conditioning group, four *S*s had the higher frequency tone as CS+.

Three other groups provided control comparisons. An *Unpaired* group received 40 tones and 40 unpaired puff presentations with the temporal scheduling identical to that of puff and CS− in the conditioning groups. Procedure for a *Two-Tone* group also retained the order and temporal intervals of trials in the conditioning groups but omitted the air puff. A sensitization control procedure with random interspersion of the UCS among both tones was not feasible since the mean intertone interval would have had to be lengthened to allow time for HR response recovery. A third control group, the *One-Tone* group, received only 40 tones on the same schedule as CS− in the paired group, and thus had an ITI ranging from 36 to 60 sec. Frequency of the tones was balanced within each group.

Instructions to all *S*s indicated that reactions "to certain distracting stimuli" were being investigated

and that heart rate, along with other responses, would be recorded. Eyelid responses were not mentioned explicitly and, as the eye blink recording headset was positioned, *S* was informed only that it would be used to record "responses." The *S* was cautioned to sit quietly and was warned that receiving credit for participation was contingent on remaining awake.

Results

Eyelid conditioning. Percentage responding to CS+ and CS− is shown in 5-trial blocks in Figure 2. Consistent with previous work in this laboratory, a CR was defined as the first pen deflection of at least 2 mm. (1 mm. eyelid closure) occurring in the interval from 200 msec. after CS onset until UCS onset. Good conditioners showed asymptotic responding to CS+ (96%) by Trial Block 5 (Trials 41–50) and achieved considerable differential responding (64%) by Trial Block 3 (Trials 21–30). Asymptotic differential responding of approximately 75% was reached by Trial Block 4. Poor *S*s, on the other hand, showed only a small increase in responding between Trial Blocks 1 and 8 and virtually no differential responding. Statistical tests over the last three trial blocks confirmed that Good *S*s responded significantly more to CS+ than to CS−, $t\ (9) = 13.96$, $p < .001$, that Poor *S*s did not respond differentially, $t\ (9) = 2.18$, $p > .05$, and that the differential responding of the Good group was significantly better than that of the Poor group, $t\ (18) = 6.88$, $p < .001$. The Good–Poor differences could not be ascribed to differing numbers of *S*s classified as voluntary responders since classification, according to the Hartman and Ross (1961) criteria, indicated that there were four voluntary responders in the Good and three in the Poor group.

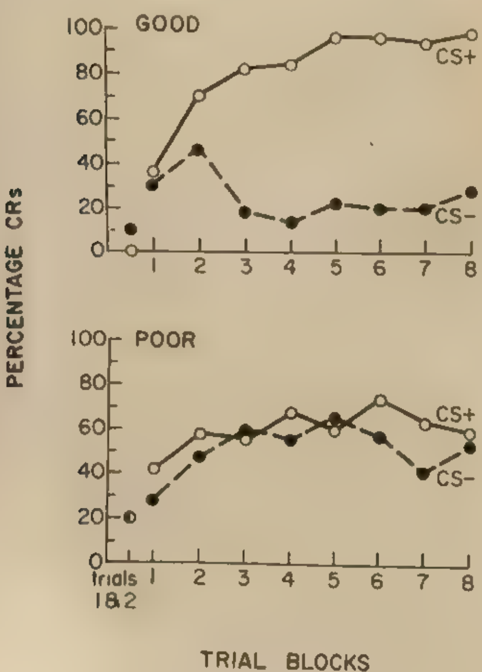


FIGURE 2. Percentage of conditioned responses for Good and Poor conditioners in blocks of five CS+ and five CS− trials.

Comparison of heart rate responses in conditioning and control groups. The differential conditioning procedure also produced highly significant HR responses which were time locked to the onset of both CS+ and CS− in both Good and Poor conditioning groups. These responses, averaged over all trials, are shown in Figure 3 for differences between HR during the first second before stimulus onset and HR on each of 6 sec. following onset. The 1-sec. averages are the bpm rate during all full and partial beats, weighted according to the fraction of the second occupied by each. A 6-sec. poststimulus period was selected for illustration and analysis since it was the longest period during which a single trend accounted for the major part of the variance and because differences from prestimulus were rarely reliable beyond this point (Putnam, 1972). Differences in prestimulus level, between groups were not significant and did not change differentially across the experimental session.

As Figure 3 illustrates, HR responses during the 6-sec. poststimulus period were generally triphasic with an initial deceleration lasting for 2 sec., a subsequent phase of acceleration also lasting for 2 sec., and a slower recovery phase. In the control groups, the accelerative phase did not always increase HR beyond prestimulus, and the response was, on the whole, less regular and less pronounced. A comparison of responses to "CS−" tones showed that

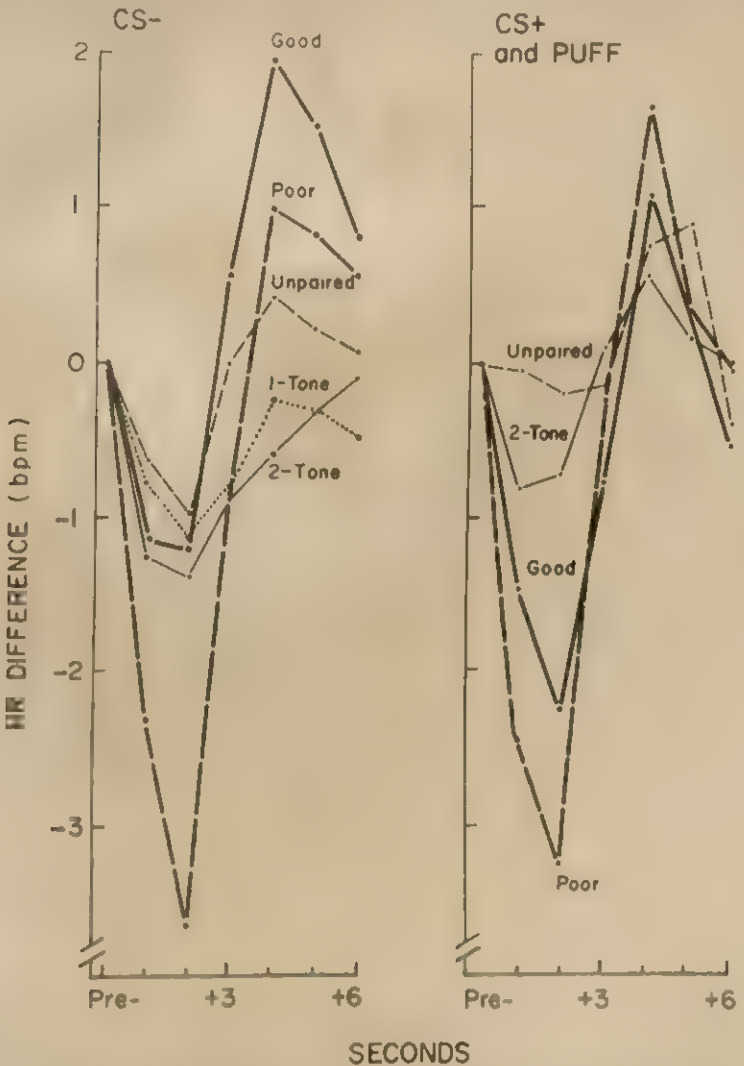


FIGURE 3. Mean heart rate (HR) responses to CS− and to CS+ alone, puff alone, or paired CS+ and puff, for each paired and control group.

differences among the five groups in overall cubic trend were significant (Table 1, Col. A) and were largely due to the paired-control comparison. The trend in each of the control groups, compared individually with the conditioning groups, was also significantly different but the controls did not differ from one another. When the three response components were tested separately by measuring changes over the first 2 seconds, over Seconds 2 to 4, and over Seconds 4 to 6, differences among groups and between paired and control groups were significant during the initial deceleration (Table 1, Col. B) and the following acceleration (Table 1, Col. C) but not during the recovery phase (Table 1, Col. D). Each of the control groups, considered separately, also differed from the conditioning groups in the degree of acceleration, but only the Unpaired group differed in initial deceleration. Again, there were no significant differences among control groups. There was also no significant dif-

ference among groups in the changes across trials, pooled into blocks of eight trials each.

The relatively large accelerations to CS in the conditioning groups were unexpected. Since the acceleration was significantly larger than in the control group given both air puff and tone unpaired, sensitization could not account for the full effect. Further, the conditioning and the unpaired groups also differed in their response to the puff. When puff was unsignaled, it elicited a significantly smaller acceleration than when it was preceded by CS+, F (1, 27) = 11.34, $p < .01$.

Comparison of heart rate responses of Good and Poor conditioners. While the HR response during conditioning differed from control group response both in initial deceleration and in the subsequent acceleration, significant differences between Good and Poor conditioners were confined to the initial 2-sec. period of deceleration which presumably reflects orienting (Table 1).

Poor conditioners were not deficient in orienting but, rather, they showed large, stable ORs which did not differ following CS+ and CS− and which did [illegible] habituate. The relative stability of [illegible] responses to CS−, as well as the consiste[illegible] group difference in magnitude of decel[illegible]n, is illustrated in Figure 4. In con[illegible] responses to CS+ (Figure 5) show[illegible] considerably more variation across tr[illegible] blocks although they were clearly less [illegible] than responses to puff in the [illegible]red group. When decelerative respo[illegible] to both CS+ and CS− and to each st[illegible]lus alone were analyzed for changes [illegible]oss blocks of eight trials, the differenc[illegible] between Good and Poor conditioners wer[illegible] reflected in a significant Groups × Stimul[illegible] × Linear Trials effect when both stimuli were included in the analysis (Table 2, Col. [illegible]) and in a significant Groups × Linear Trials effect for response to CS+ (Table 2, Col. B). Only the Groups effect was significant in response to CS− (Table 1, Col. C).

The finding of greatest interest was [illegible] change in the decelerations elicited [illegible] CS+. Both Good and Poor conditio[illegible]s showed a similar increase in deceler[illegible]n over the first three trial blocks but, on [illegible]er

TABLE 1

TESTS OF DIFFERENCES IN THE HEART RATE RESPONSE OF TWO PAIRED (PR[illegible] AND THREE CONTROL (C) GROUPS TO CS[illegible]

Source of variation	df	A. Cubic trend	B. Initial deceleration	C. Acceleration	D. Recovery
		F ratios			
Groups (G)	4	7.4**	4.1**	9.5**	1.2
Pr–C	1	26.2**	6.3*	33.5**	2.4
Among C	2	.1	.3	.6	1.5
Good–bad	1	3.2	9.9**	4.0	.9
Nonorthogonal					
Group comparisons (cfs)					
Pr–unpaired	1	13.2**	4.6*	14.4**	.4
Pr–1 tone	1	15.3**	3.7	20.4**	.7
Pr–2 tone	1	15.3**	2.4	21.5**	3.9
Trials (T)	4	.5	.6	1.0	.3
Linear T	1	.3	0.0	.8	.2
G × T	16	1.2	1.4	1.3	.9
G cfs × Linear T	1–4	*ns*	*ns*	*ns*	*ns*
		Error mean squares			
Ss/G	45	38.7	14.8	14.3	12.8
Ss/G × T	180	11.4	4.2	4.5	3.8
Ss/G × Linear T	45	12.5	3.4	5.4	5.0

Note. Abbreviation: CS− = negative conditioned stimulus.
* $p < .05$.
** $p < .01$.

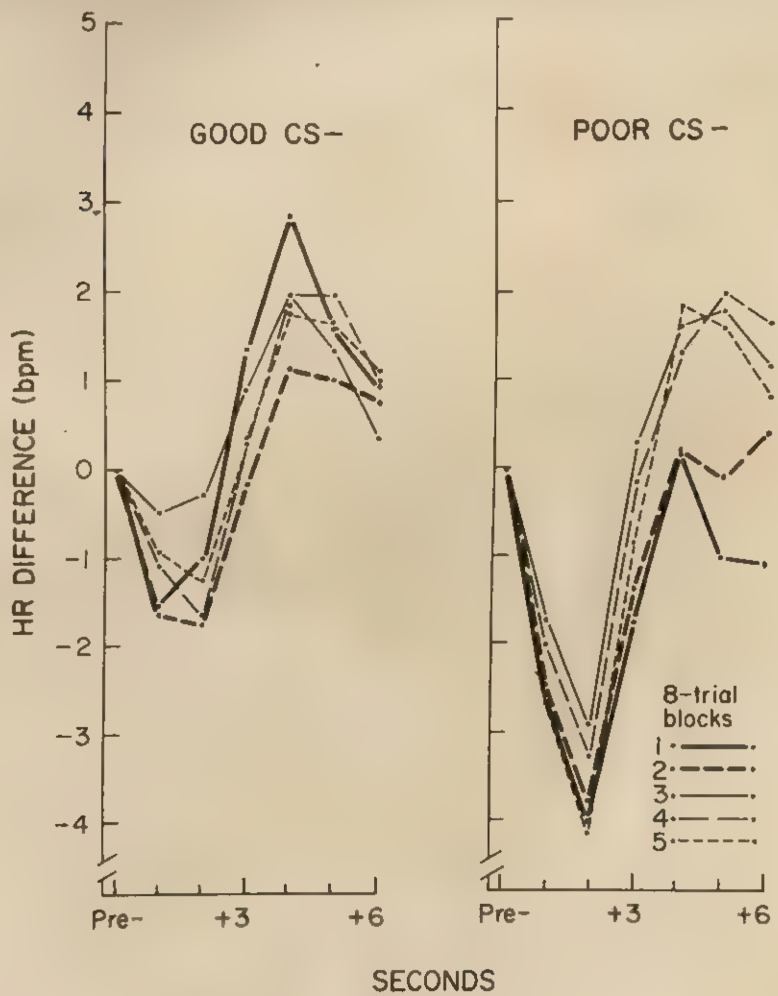


FIGURE 4. Heart rate (HR) responses of Good and Poor conditioners to CS− in five trial blocks of eight trials each.

trials, Poor conditioners continued to give large decelerations while the response showed habituation in the Good conditioning group. These changes can be seen more clearly in Figure 6, which also shows the average decelerative response of those Poor conditioners who avoided the puff on at least 18 of the final 20 CS+ trials by giving a CR coincident with the UCS. Although there were only four subjects in this category, their failure to show any habituation seems to rule out the possibility that ORs of Good conditioners habituated because they were able to avoid the puff. The temporal correspondence between orienting and conditioning is illustrated in Figure 7 which shows the differential eyelid responding of Good conditioners superimposed on their HR deceleration data. It is evident that the period of maximum deceleration corresponds roughly to the time when differential eyelid conditioning reached an asymptote.

For comparison with studies that have described the learning behavior of *S*s categorized as high and low orienters on the basis of some measure of initial orienting, such as response to the first CS, differential conditioning scores of all *S*s were correlated with deceleration to the first

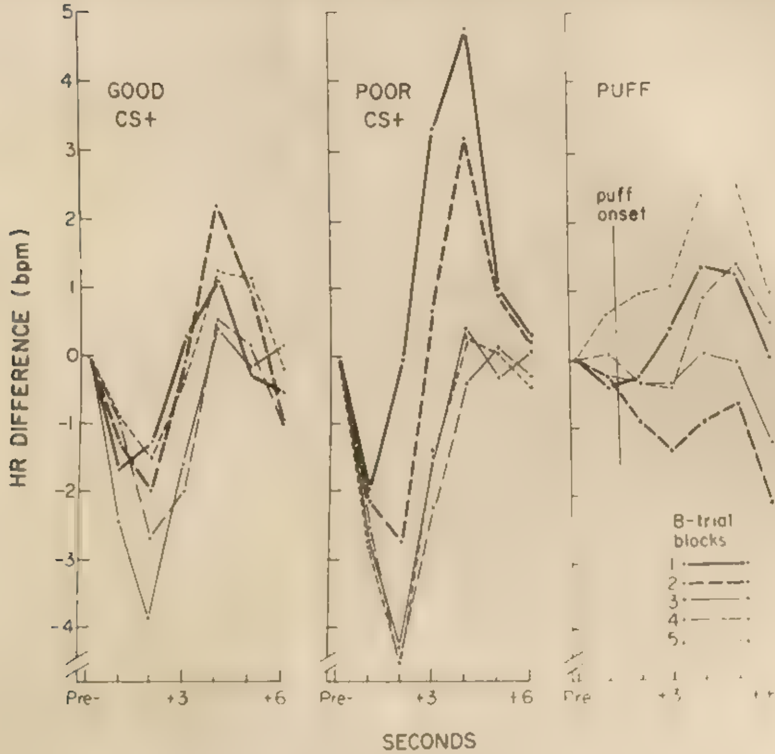


FIGURE 5. Heart rate (HR) responses of Good and Poor conditioners to CS+ and of Unpaired *S*s to air puff in five trial blocks of eight trials each.

CS− and to the first CS+. Neither correlation was significant and they differed in direction (Pearson product-moment $r = -.30$ for first CS− and $r = +.36$ for first CS+).

Discussion

The present study of the relation between orienting, measured by HR deceleration, and learning, measured by acquisition of a differentially conditioned blink response, confirms the descriptions of Soviet investigators (e.g., Sokolov, 1963; Vinogradova, 1965). The OR intensified during acquisition and habituated as conditioning stabilized. Similar results have also been reported in a recent study of HR changes accompanying conditioning of the rabbit's corneoretinal potential (Powell, 1972). Since the Soviet studies providing the best empirical basis for the OR-learning relation either conditioned an autonomic response and thus confounded OR and CR measures, or employed an instructed motor response and thus might have imposed attentional requirements specific to a "voluntary" task, the present results place on a firmer basis the generalization that during learning first increases and then declines while a response increases and stabilizes. rise and fall during learning is also of inter because it suggests that the common finding of similar inverted-U functions in autono conditioning studies may reflect changes in orienting rather than in a learning process

Two other findings of the present study that *S*s failing to condition had large, stabl ORs and that HR accelerations to the UCS were larger to a signaled than to an unsignaled UCS, stand in contrast to most previous work. The existence of larger ORs in Poor than in Good conditioners, both on the first trial and throughout conditioning, is compatible with Soviet versions of OR theory which emphasize that very marked orienting, as well as very weak, interferes with conditioning, but the finding does not agree with Western versions that propose only a positive relation between orienting and conditioning (e.g., Maltzman, 1967). Maltzman views the evidence of a positive relation and the absence of evidence of any interfering effects of high OR on learning as important for the argument that orient-

TABLE 2

TESTS OF DIFFERENCES IN THE DECELERATIVE HEART RATE RESPONSES OF GOOD AND POOR CONDITIONERS

Source of variation	*df*	A. CS+ and CS−	B. CS+	C. CS−
		F ratios		
Groups (G)	1	3.4	.8	6.4*
Stimuli (S)	1	.7	—	—
Trials (T)	4	3.1*	6.5**	1.0
Linear T	1	7.3*	9.5**	0.0
Quadratic T	1	4.8*	20.5**	.9
G × T	4	1.3	2.3	.4
G × Linear T	1	3.8	6.0*	.1
G × S	1	3.4	—	—
S × T	4	4.6**	—	—
S × Linear T	1	6.4*		
S × Quadratic T	1	9.0**		
G × S × T	4	1.5	—	—
G × S × Linear T	1	4.7*		
		Error mean squares		
Ss/G	18	42.2	27.2	22.8
Ss/G × S	18	7.8	—	—
Ss/G × T	72	4.4	5.5	4.6
Ss/G × Linear T	18	5.1	7.5	3.0
Ss/G × Quadratic T	18	3.4	3.2	6.3
Ss/G × S × T	72	5.7	—	—
Ss/G × S × Linear T	18	5.4	—	—

Note. Abbreviations: CS+ = positive conditioned stimulus, CS− = negative conditioned stimulus.
* $p < .05$.
** $p < .01$.

ing responses are not measuring a generalized arousal or drive state. The argument is reasonable but it is questionable that the previously existing empirical data permit a distinction between arousal and OR effects on learning.

A major flaw in the evidence for a solely positive OR-learning relation not due to arousal is that a number of studies have classified *Ss* into high- and low-OR groups on the basis of only a single trial which has sometimes been the first habituation and sometimes the first UCS trial (e.g., Maltzman, 1967). As Zeiner and Schell (1971) point out, the response to a noxious UCS qualifies as a DR. Base levels of skin conductance have also been employed (Maltzman & Mandell, 1968) as a measure of the tonic OR but might equally well be considered a measure of arousal (Malmo, 1959). A second major problem with the evidence is that studies have not adequately controlled for general reactivity effects. Many have measured CR acquisition,

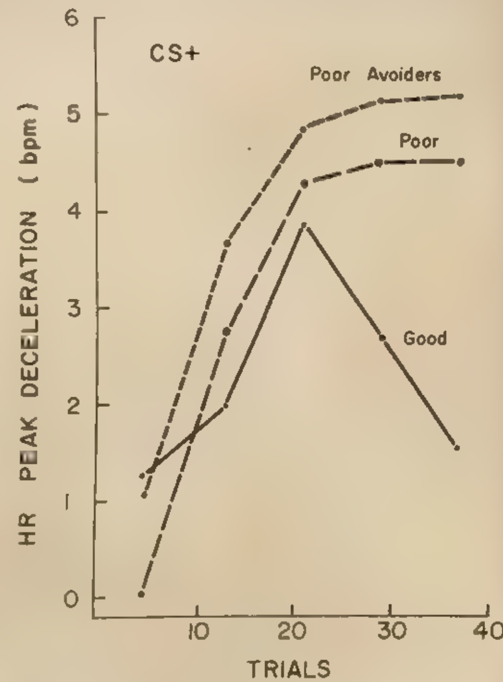


FIGURE 6. Changes over CS+ trials in the peak heart rate (HR) deceleration (prestimulus HR less Second-2 HR) of Good conditioners, Poor conditioners, and Poor avoiders.

which is more likely than differential conditioning to reflect arousal. Even differential conditioning may not rule out activation effects if OR and CR are measured by the same re-

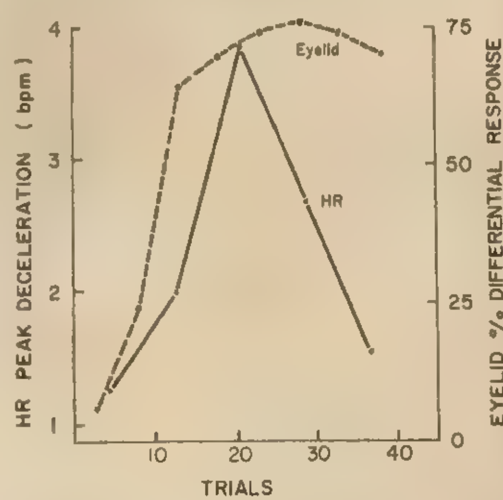


FIGURE 7. Change over trials in the relation between differential eyelid responding of Good conditioners and their peak heart rate (HR) decelerations to CS+.

sponse system and there are *Ss* who fail to respond to both CS+ and CS− in a given trial block. In such cases, there is no measure of whether discrimination is or is not present and their inclusion as instances of no discrimination artifactually lowers the discrimination score. Since low-OR *Ss* are more likely, on the basis of a response reactivity factor, to have no-response trials, reactivity could thus affect the OR-differential conditioning relation as well as the relation of orienting to CR acquisition. Zeiner and Schell (1971) provide an example of this in showing that a positive relation between a GSR-OR and differential conditioning of first-interval GSR responses disappeared when discrimination was measured only on trials yielding some nonzero response.

While much, if not all, of the data purportedly showing a positive OR-learning relation may thus reflect the effects of a DR, arousal, or specific response reactivity rather than an OR, these effects could not have produced the OR-learning relation of the present study since the OR was measured by HR deceleration, which is *not* a measure of arousal, and learning was measured by the differential conditioning of a response in a different system. However, findings of a significant negative relation between the OR and learning appear to be nonexistent in the Western literature, except for the present study, and it is possible that there are critically important differences between how orienting affects processes developing during conditioning with a short as opposed to a long CS–UCS interval.

Paradigm differences may also be important in accounting for the second discrepant finding of the present experiment, that HR accelerations were larger in response to a signaled than to an unsignaled UCS. A number of studies, generally measuring GSR, have found that unconditioned response (UCR) amplitude is smaller, not larger, in response to a signaled UCS (e.g., Kimble & Ost, 1961; Kimmel, 1967; Peeke & Grings, 1968). Since one interpretation of UCR diminution is that a response to a signaled UCS "is due to a carryover of the response processes begun during the inter-stimulus interval [Grings, 1969, p. 607]," it might be thought that while a unidirectional response like GSR would be interfered with by a prior response, a bidirectional system like HR might be enhanced, if only by virtue of an initial level effect. A recent report of UCR diminution with HR as the response measure (Lykken, MacIndoe, & Tellegen, 1972) makes it unlikely that the explanation lies solely with GSR–HR differences, but an alternative is suggested by considering differences in the length of the CS–UCS interval.

The Lykken et al. (1972) study, in common with GSR studies of UCR diminution and many OR-learning studies, employed a shock UCS and a relatively long CS–UCS interval, initiated 5 sec. before shock onset by the first of two successive CSs. In contrast, the present study employed a CS–UCS interval of only 800 msec. which would require that any preparatory processes be mobilized very quickly. Just what this implies for HR changes is problematic but there is evidence that the cardiovascular system participates both in orienting and in preparations for action and that preparations for action result in HR acceleration (e.g., Chase, Graham, & Graham, 1968). Since HR reflects only one aspect of cardiovascular preparations, is slow relative to skeletal responses, and shows lag as a result of feedback to the heart from blood pressure, volume, and O_2 changes, HR changes associated with mobilizing for rapid action would be expected to outlast the action and thus might summate with the UCR. This hypothesis, of an interaction between length of the CS–UCS interval and the type of UCR measured, receives support from another recent study (Furedy & Klajner, 1972) which also employed a short CS–UCS interval (500 msec.) and also obtained a larger UCR to a signaled UCS when the response was a cardiovascular change, peripheral vasomotor constriction. There was no signaling effect when GSR was measured.

The argument, then, agrees with Grings (1969) that the signaled UCR reflects a carry-over of response processes begun with the CS but suggests additionally that whether the UCR is increased or decreased depends upon the length of the CS–UCS interval and the particular response that is measured. These same factors may also determine whether an orienting response, with which brief suspension of movement is associated (Obrist, Webb, & Sutterer, 1969) interferes with or facilitates the occurrence of a particular CR. With short CS–UCS intervals, a prolonged, pronounced OR may not be required in order to establish an association between two stimuli in close temporal proximity and, if it does occur, might be expected to interfere with execution of a skeletal response such as blink. With longer CS–UCS intervals, prolonged orienting may be more important in establishing an

association between stimuli more widely separated in time and even if it were very pronounced would not interfere with an autonomic CR which was itself part of the orienting complex.

REFERENCES

Chase, W. G., Graham, F. K., & Graham, D. T. Components of heart rate response in anticipation of reaction time and exercise tasks. *Journal of Experimental Psychology*, 1968, **76**, 642–648.

Furedy, J. J., & Klajner, F. Unconfounded autonomic indexes of the aversiveness of signaled and unsignaled shocks. *Journal of Experimental Psychology*, 1972, **92**, 313–318.

Graham, F. K. Habituation and dishabituation of responses innervated by the autonomic nervous system. In H. V. S. Peeke & M. J. Herz (Eds.), *Habituation: Behavioral studies and physiological substrates.* Vol. 1. New York: Academic Press, 1973. Pp. 163–218.

Grings, W. W. Anticipatory and preparatory electrodermal behavior in paired stimulation situations. *Psychophysiology*, 1969, **5**, 597–611.

Hartman, T. F., & Ross, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *Journal of Experimental Psychology*, 1961, **61**, 334–338.

Jackson, J. C. Elicitation and habituation of the orienting reflex at differing levels of stimulus intensity. Unpublished doctoral dissertation, University of Wisconsin, 1972.

Kimble, G. A., & Ost, J. W. P. A conditioned inhibitory process in eyelid conditioning. *Journal of Experimental Psychology*, 1961, **61**, 150–156.

Kimmel, E. Judgments of UCS intensity and diminution of the UCR in classical GSR conditioning. *Journal of Experimental Psychology*, 1967, **73**, 532–543.

Lykken, D. T., MacIndoe, I., & Tellegen, A. Preception: Autonomic response to shock as a function of predictability in time and locus. *Psychophysiology*, 1972, **9**, 318–333.

Lynn, R. *Attention, arousal, and the orientation reaction.* Oxford, England: Pergamon Press, 1966.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychological Review*, 1959, **66**, 367–386.

Maltzman, I. Individual differences in "attention": The orienting reflex. In R. M. Gagné (Ed.), *Learning and individual differences.* Columbus, Ohio: Charles E. Merrill, 1967. Pp. 94–112.

Maltzman, I., & Mandell, M. P. The orienting reflex as a predictor of learning and performance. *Journal of Experimental Research in Personality*, 1968, **3**, 99–106.

Morgenson, D. F., & Martin, I. The orienting response as a predictor of autonomic conditioning. *Journal of Experimental Research in Personality*, 1968, **3**, 89–98.

Obrist, P. A., Webb, R. A., & Sutterer, J. R. Heart rate and somatic changes during aversive conditioning and a simple reaction time task. *Psychophysiology*, 1969, **5**, 696–723.

Peeke, S. C., & Grings, W. W. Magnitude of UCR as a function of variability in the CS–UCS relationship. *Journal of Experimental Psychology*, 1968, **77**, 64–69.

Powell, D. A. Autonomic-somatic relationships: Heart rate and eyeblink conditioning in the rabbit. Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1972.

Prokasy, W. F. Do *D* and *H* multiply to determine performance in human conditioning? *Psychological Bulletin*, 1967, **67**, 368–377.

Putnam, L. E. Heart rate correlates of the classically conditioned eyelid response. Unpublished master's thesis, University of Wisconsin, 1972.

Razran, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychological Review*, 1961, **68**, 81–147.

Sokolov, E. N. Neuronal models and the orienting reflex. In M. A. B. Brazier (Ed.), *The central nervous system and behavior.* New York: Josiah Macy Jr. Foundation, 1960. Pp. 187–276.

Sokolov, E. N. *Perception and the conditioned reflex.* New York: Pergamon Press, 1963.

Vinogradova, O. S. On the dynamics of the orienting reflex in the course of closure of a conditioned connection. In L. G. Voronin, A. N. Leontiev, A. R. Luria, E. N. Sokolov, & O. S. Vinogradova (Eds.), *Orienting reflex and exploratory behavior.* Washington, D.C.: American Institute of Biological Sciences, 1965. Pp. 45–53.

Wilcox, S. M., & Ross, L. E. Differential classical eyelid conditioning as a function of CS intensity, CS rise time, and interstimulus interval. *Journal of Experimental Psychology*, 1969, **82**, 272–278.

Zeiner, A. R., & Schell, A. M. Individual differences in orienting, conditionability, and skin resistance responsivity. *Psychophysiology*, 1971, **8**, 612–622.

(Received April 23, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 574–578

PRAGMATIC NORMALIZATION:

FURTHER RESULTS FOR SOME CONJUNCTIVE AND DISJUNCTIVE SENTENCES[1]

SAMUEL FILLENBAUM[2]

University of North Carolina, Chapel Hill

On a paraphrase task involving pragmatically extraordinary sentences *S*s normalized two sorts of disordered conjunctive sentences as well as disjunctive sentences involving "perverse" threats more than half the time. Later when comparing their paraphrases with the source sentences they were often unable to detect differences in meaning. These results were interpreted in terms of a basic assumption made by many *S*s that messages are sensible, that what is described in messages must conform to the customary order and connection among events, and that apparent anomalies can be attributed to inadequacies in the encoding of the messages.

The tendency to normalize connected discourse or stories in memory has been well known since at least the time of Bartlett (1932). It may be the case that such normalization involves not so much or not only memorial processes, but may largely be a consequence of the way in which strange material is initially interpreted and stored, such that extraordinary descriptions are transmuted into ones more mundane and commonplace, given knowledge of the ways things usually transpire, and a strong presumption of the sensibleness of the material encountered. The present research constitutes a replication and extension of part of an earlier study (Fillenbaum, 1971) in which it was found that on a paraphrase task verb order was often permuted for disordered conjunctive sentences, normalizing them and assimilating the order of events described to conventional order. It is an extension in two regards: (*a*) in that (*i*) a new kind of ordered conjunctive sentence is examined in addition to the kind studied before, and (*ii*) a class of strange disjunctive sentences ("perverse" threats) is also studied, and, perhaps more important, (*b*) in that we also s[illegible] determine whether *S* could detect [illegible] changes in his paraphrases, by havin[illegible] review each of his paraphrases at the en[illegible] and having him say whether there was any difference in meaning between the original sentence and his paraphrase, and, if so, to indicate what sort of difference was involved.

Method

There were 40 sentences to be paraphrased. There were 5 filler items at the start and then, in randomized order (*a*) 10 ordered conjunctive sentences in normal order (e.g., "John got off the bus and went into the store"), (*b*) 10 ordered conjunctive sentences with normal order reversed, these sentences being extraordinary but semantically coherent (e.g., "John dressed and had a bath"), (*c*) 4 ordered conjunctive sentences with normal order reversed, these sentences violating temporal entailment relations between their main verbs (e.g., "John finished and wrote the article on the week end"), (*d*) 5 disjunctive sentences involving commonplace threats (e.g., "Stop that noise or I'll call the police"), (*e*) 5 disjunctive sentences involving conditional threats that were "perverse" in that the consequent clause indicated that the speaker would *not* bring about some undesirable event commonly to be expected in the circumstances (e.g., "Don't print that or I won't sue you"), plus a single "perverse" warning, the sentence "Get a move on or you will catch the bus."

Each of the 51 *S*s, who were college students meeting a class requirement for an introductory psychology course, was run individually being given a deck of 3 × 5 in. cards each containing one of the sentences to be paraphrased, and a blank deck for writing down his paraphrases. The instructions, which were the same as the version two (strong) instructions used in the earlier study (Fillenbaum,

[1] This work was supported in part by U.S. Public Health Service Research Grant MH-10006 from the National Institute of Mental Health. The writer is indebted to R. Hall for his aid in gathering and analyzing the data.

[2] Requests for reprints should be sent to Samuel Fillenbaum, University of North Carolina, Psychometric Laboratory, Chapel Hill, North Carolina 27514.

1971), stressed that the task was to paraphrase or rephrase "the sentences as accurately as you can, conserving meaning as completely as possible" and that *S* was not "to improve the sentences or make them more sensible, but to *paraphrase* them, rewording each in a way that captures its meaning as accurately as possible." After *S* had finished the paraphrase task at his own pace he was told that he would have to go through his paraphrases, that he would be asked "about the remaining differences IF ANY" between each sentence and its paraphrase, and that if he did "see some shred of difference" to say what seemed to be involved. For each item *S* was then shown the original sentence and his paraphrase and asked "Is there any difference in meaning between the sentence and your rewording?" and if this question was answered "yes," *S* was asked "What sort of difference?" After running through all the items in this fashion if *S* had actually changed the meaning of the sentences in his paraphrases and noticed this at all during the review of the items (which was the case for most *S*s), he was confronted with this fact, reminded of the instructions to preserve meaning, and asked "How come?"

The analysis of paraphrases was relatively straightforward. For each conjunctive sentence one could determine whether or not the order of verb phrases had been maintained. Change in the order of verb phrases for the disordered conjunctive sentences almost always resulted in normalizing the sentences. In addition there were occasional paraphrases which did not change the unusual order of verbs but which elaborated or emended the sentences in such fashion as to turn them into sensible, normal sentences. The disjunctive sentences, threats, were paraphrased in various ways perhaps most commonly as conditionals. One could generally unambiguously determine whether or not these paraphrases normalized meaning. Thus, for example, given "Clean up the mess or I won't report you," both (*a*) "If you don't clean up the mess I'll report you" and (*b*) "If you clean up the mess I won't report you" may be regarded as changing and normalizing the original sentence, while both (*c*) "If you clean up the mess I'll report you" and (*d*) "If you don't clean up the mess I won't report you" may be regarded as conserving some of the strange properties of the original sentence.

The judgments *S*s made on reviewing their paraphrases, as to whether or not these differed in meaning from the original sentences, could be coded simply as "different" or "same." The observations *S*s made when they did judge a paraphrase to be different from its source sentence, and the various things said by way of explanation or self-justification at the end, when confronted with the fact that their paraphrases had often involved meaning changes, were found to be codeable in a limited number of categories (see results below).

Results and Discussion

For normally ordered conjunctive sentences meaning was preserved in 99% of the paraphrases (with 90% judged same in meaning and 9% different in meaning) and was perverted in 1% of the paraphrases (with all of these judged different in meaning). For disordered conjunctive sentences violating a contingent relation meaning was preserved in 36% of the paraphrases (with 31% judged same in meaning and 5% different in meaning) and normalized in 64% of the paraphrases (with 22% judged same in meaning and 42% different in meaning). For disordered conjunctive sentences violating an entailment relation meaning was preserved in 30% of the paraphrases (with 28% judged same in meaning and 2% different in meaning) and normalized in 70% of the paraphrases (with 43% judged same in meaning and 27% different in meaning). For normal conditional threats meaning was preserved in 99% of the paraphrases (with 95% judged same in meaning and 4% different in meaning) and perverted in 1% of the paraphrases (with all of these judged different in meaning). For "perverse" conditional threats meaning was preserved in 46% of the paraphrases (with 40% judged same in meaning and 6% judged different in meaning) and normalized in 54% of the paraphrases (with 34% judged same in meaning and 20% different in meaning).[3]

It is clear that ordered conjunctive sentences in normal order and normal conditional threats are hardly ever perversely paraphrased, and that in each case paraphrases that preserve meaning are overwhelmingly regarded as doing just that. For these sorts of sentences when *S*s claim that paraphrases change meaning the most common ground given is that change in a particular single word led to a change in sentence meaning. Results for the two sorts of disordered conjunctive sentences

[3] These results, insofar as comparable, are very similar to those obtained earlier (Fillenbaum, 1971) where meaning was preserved in 96% of the paraphrases of normally ordered conjunctive sentences, but for only 39% of the paraphrases of disordered conjunctive sentences violating a contingent relation. In a previous (unpublished) study it was found that paraphrases preserved meaning for 99% of normal threats, but for only 43% of "perverse" threats.

and for the "perverse" conditional threats are quite different. For the disordered conjunctive sentences 64% (contingent relation violated) and 70% (entailment relation violated) of the paraphrases involve normalization, while for the "perverse" conditional threats 54% of the paraphrases involve normalization (the results for the single "perverse" warning are roughly similar to those for "perverse" threats and will not be mentioned further). The differences in normalization for the three kinds of strange sentences do not reach significance when proportions of normalization of each sentence kind are contrasted for each S by means of sign tests.

Turning next to judgments of whether or not paraphrases changed the meaning of the various kinds of strange sentences, we find that for disordered conjunctive sentences where a contingent relation between verbs has been violated normalizing paraphrases are detected almost twice as often as not (the figures are 42% and 22%), while for disordered conjunctive sentences where an entailment relation between verbs has been violated normalizing paraphrases are detected less often than they are overlooked (the figures are 27% and 43%). The difference in detection rates is significant ($p < .01$) when proportions of detection of normalizing changes for the two kinds of sentences are contrasted for each S by means of a sign test, indicating that Ss are responsive to differences in the semantic properties of these two kinds of disordered sentences even though the actual percentages of normalizing paraphrases are quite similar (64% and 70%). For "perverse" threats we find that paraphrases that normalize the source sentences are detected less often than they are overlooked (the figures are 20% and 34%), and that the detection rate for these normalizing changes is significantly less than that for normalizing changes for disordered conjunctive sentences violating a contingent relation ($p < .05$), and no different from that for normalizing changes for disordered conjunctive sentences violating an entailment relation. Given considerable conceptual differences between the disordered conjunctive sentences and the strange disjunctive sentences it is difficult to say anything as to the interpretation of differences in detection rate for normalizing changes.

Consider next what Ss say is involved when they do detect differences between strange source sentences and their paraphrases of them. For disordered conjunctive sentences of the two kinds studied we find that the great majority of comments refer to order change, change in temporal sequence, and change in causal sequence, with 85% of the comments for sentences where a contingent relation is violated and 65% of the comments for sentences where an entailment relation is violated falling into these three categories. These comments are quite realistic for it is just these sorts of changes that have occurred in the meaning normalizing paraphrases. If we consider the comments made when paraphrases of "perverse" threats are judged to be different from the source sentences, we again find some use of the above three categories but now the most common observation is that the paraphrase is a sentence opposite in meaning to its source. Again such comments are quite realistic, for normalization of "perverse" threats characteristically involves just that, see, e.g., the difference between the source sentence "Clean up the mess or I won't report you" and a paraphrase such as "If you clean up the mess I won't report you."

The results just presented raise two questions: (*a*) If Ss can detect differences between their paraphrases and the source sentences with regard to such significant matters as changes in temporal and causal order and changes which result in opposite meaning, then why did they provide such misleading paraphrases, given strong instructions to preserve and conserve meaning and warnings against attempting to improve or make the sentences more sensible? (*b*) If Ss cannot detect such substantial differences between the source sentences and their paraphrases then what might be responsible for such obtuseness or blindness? What is empirically true, and presupposed in both of the above questions, is the fact that extraordinary sen-

tences are often pragmatically normalized in the paraphrases, and one needs to ask (*c*) why this should be so.

It seems plausible that the same general answer holds for all three questions. Even in the peculiar circumstances of the psychological laboratory *S*s seem to be acting on the basic assumption that *what is described in discourse will be sensible*, that what is described will conform to the customary order of events and will satisfy normal qualitative and causal relations between events or actions. If a sentence is encountered which appears to violate this assumption one may consider it to be a clumsy, inadequate phrasing or description, and turn it into a more appropriate version in one's paraphrase. Thus what is taken as awry or extraordinary is not the world, but the linguistic account of it. As to why *S*s should have provided normalizing paraphrases when they *could* detect significant differences between these and the source sentences, it may be that as far as *S*s are concerned a difference detected is not so much one between descriptions of two different sorts of events as one between two different descriptions of the same event, with the paraphrase expressing properly what is *intended and badly expressed* by the original sentence. Finally, why might *S*s have failed to notice differences between the source sentences and their normalizing paraphrases? This may just constitute further testimony to the strength of the basic assumption, such that the original "malphrased" sentences and their own "improved" versions are really regarded as amounting to the same description of sensibly ordered and organized events.

The above suggestions while speculative find support in the comments *S*s made at the end, by way of explanation and justification for their paraphrases. Of the 51 *S*s 10 or 11 hardly ever changed meaning in their paraphrases, so we need to consider only the responses of the other 40 *S*s.[4] Comments which claimed (*a*) that the paraphrases made things clear and more sensible, (*b*) that the paraphrases put things into natural order, (*c*) that the original sentences violated expectancies, (*d*) that the original sentences were illogical, and (*e*) that they knew what the original sentences were trying to say and so they said it, account for the responses of 25 of these *S*s. All of them appear to be claiming, in one way or another, that their paraphrases somehow put more adequately what was badly phrased in the original sentences. In addition there were 7 *S*s who said that they misread or read incorrectly the original sentences, which suggests, if their comments are taken at face value, that they may not even have noticed the strange properties of the original sentences, and that in the very act of reading and coping with them they corrected or normalized these sentences. Thus for 32 of the 40 *S*s there is some evidence for the operation of the assumption that what is described in discourse must be sensibly organized. One further observation is worth reporting, almost every one of the 40 *S*s who was trying to explain or justify his (normalizing) paraphrases seemed in part inarticulate, and was considerably flustered. It seemed as though it was somehow very important for *S*s to adjust things to conventional order and sequence, and to exhibit the normal relations among events, so as to say what the speaker of the source sentence must have intended or meant.

People may make the basic assumption that sentences are sensible just because they believe that the events of which such sentences purport to be accounts occur or are organized in characteristic ways, and that it is much more likely that the world that is being described is the commonplace one and that the description is burdened with the errors and malapropisms of the speaker than that extraordinary events are being

[4] In the earlier study there were consistent individual differences, with some *S*s normalizing the majority of the disordered sentences and others hardly normalizing any (Fillenbaum, 1971). In the present study similar results were obtained both with regard to whether or not strange source sentences were normalized and with regard to detection of such changes.

accurately and faithfully presented. It is as though people focus not on linguistic messages per se, but on the information they embody or appear to convey, considering and assimilating this information in relation to their preexisting knowledge of the ways of the world.

REFERENCES

Bartlett, F. C. *Remembering.* Cambridge, England: Cambridge University Press, 1932.

Fillenbaum, S. On coping with ordered and unordered conjunctive sentences. *Journal of Experimental Psychology,* 1971, **87,** 93–98.

(Received March 24, 1973)

MODALITY DIFFERENCES IN RECOGNITION MEMORY FOR WORDS AND THEIR ATTRIBUTES

KIM KIRSNER[1]

University of Western Australia, Nedlands, Australia

[illegible] Experiment I, 20 Ss gave recognition responses for which accuracy and reaction time data were collected. In Experiment II, 20 new Ss gave [illegible] items [illegible]. Intramodality recognition performance was superior to [illegible] performance [illegible] modality report data [illegible] The results were thought to be consistent with a [illegible] explanation of recognition facilitation and the modality report data.

When a person hears or sees a word, one possible assumption is that the word is immediately translated into its semantic form. That is, the meaning of the stimulus is extracted and retained while normally redundant information about the physical characteristics of the stimulus are lost. From the viewpoint of a model of memory, it is of interest that information concerning the physical nature of the stimulus may be retained for a period of time apparently beyond that which is necessary for semantic analysis. The present study was implemented with a view to (*a*) establishing the longevity of this representational storage and (*b*) gaining some insight into the coding processes involved in the retention of the representational as distinct from semantic attributes of verbal stimuli.

The present study used a modification [illegible] Feghtenman (1961). In their procedure, Ss were exposed to a series of words within which each item was repeated once, at one of several possible retention intervals. The technique was modified here so that an item could be repeated in the same modality (e.g., visual visual—VV) or in the alternative modality (e.g., auditory visual—AV) on the 2 presentations. Two distinct questions are considered. The first question concerns the duration of persistence of modality-specific information in memory. In the present paradigm this may be inferred from the duration of intramodality facilitation in recognition performance, for example, from the presence of an advantage for VV relative to AV. The second question follows Mann (1970) in adopting a direct approach in which S is required to [illegible] modality of presentation of each item recognized as *old*. The distinction between recognition facilitation and knowledge about modality of presentation may be important in that, in one case, the information may facilitate recognition even though it is not directly accessible, whereas, in the other case, the information may be directly accessible without benefiting recognition performance per se.

In experiments that have used words as stimuli, research has shown that auditory [illegible]tation is associated with better performance than visual presentation over a variety of memory tasks. However, the duration of this auditory superiority appears to be technique dependent, ranging

[1] [illegible]

from the last 2–3 serial positions in free recall (Murdock & Walker, 1969) to as many as 7–10 serial positions or seconds in probe recall (Murdock, 1967) and recognition memory (Kirsner & Craik, 1971). As the auditory advantage was found over all sampled serial positions in the probe recall and recognition experiments, the present study was designed to sample recognition performance over a substantially longer set of serial positions.

The position with regard to visual persistence for linguistic stimuli is complicated. On one hand, it has been widely assumed, following Sperling (1967), that apart from transient retention in an iconic form there is little or no visual persistence for linguistic stimuli. In line with this conclusion a number of mixed-modality studies using discrete-trial recognition and probe recall procedures have found no evidence for short-term visual persistence (Burrows, 1972; Kirsner & Craik, 1971; Murdock, 1967). Against this, however, results from clinical and shadowing studies using recall as well as *same different* techniques have found evidence that visual information for 1–3 letter stimuli persists for up to 25 sec. (Kroll, Parks, Parkinson, Bieber, & Johnson, 1970; Parkinson, 1972; Parks, Kroll, Salzberg, & Parkinson, 1972; Warrington & Shallice, 1969). One possible explanation of the inconsistency may be that visual persistence is masked in mixed-modality studies by a short-term auditory component. If this is so, then a longer sampling period may be required in order to reveal visual facilitation effects. This condition is met in the present study, wherein performance was sampled over a range of 31 intervening items in the continuous list.

The purpose of the experiments reported here was to explore the duration and characteristics of representational persistence in audition and vision. Experiment I compared performance, as measured by both accuracy and recognition reaction time (RT), for intra- and intermodality recognition decisions. Experiment II also provided data on recognition accuracy and, in addition, required *Ss* to report the modality of first presentation for recognized items.

Method

Subjects. Twenty undergraduate *Ss* from the University of Toronto *S* pool were used in each experiment. They were naive with respect to the experiment and were paid for their services on completion of the single session. All *Ss* were tested individually.

Design. The experiments were conducted using a within-*Ss* design for the 2 × 2 × 6 factorial combination representing test modality (auditory or visual), similarity (same or different modality on first and test presentations), and lag (1, 2, 4, 8, 16 and 32). Lag = 1 refers to an immediate repetition with no intervening items.

The *Ss* were presented with a continuous series of 772 items. The first 52 items were intended to absorb short-run practice effects and were not considered in the analyses. The 720 critical items comprised 360 words presented twice each, with 15 replications within each of the 24 experimental conditions. Each replication of the 24 *test* trials was completed within a concealed block of 48 consecutive trials, and the experimental conditions were ordered differently within each replication.

The words, drawn from a pool of common, 2-syllable nouns of 5, 6, 7, or 8 letters, were presented at a rate of 4 sec/word. Visual items were exposed for .5 sec. and were followed by a blank interval for the remaining 3.5 sec. of the intertrial interval. This was followed immediately by an auditory item or another visual item in accordance with the predetermined sequence. Two formats were drawn up, each format constituting a random sample from the word pool. Each format was presented to 10 *Ss* in each experiment.

Procedure. The visual items were prepared in lowercase on an IBM Selectric typewriter. For presentation, the list was inserted in the typewriter; the list was advanced by a solenoid attached to the return button. A Shibadan television camera was focused on the return carriage, and *Ss* viewed the stimuli on a Shibadan monitor. The typewriter was advanced twice for each visual word, once to bring the item onto the screen and once to introduce the blank phase. The screen remained blank during presentation of the auditory words.

Auditory items, previously recorded in a female voice at intervals appropriate to the mixed-modality sequence were presented through a loudspeaker located immediately above the television screen from one track of a stereo tape recorder. The auditory and visual sequences were synchronized by feeding tones, located for the visual sequence, from the second auditory track through a speech relay to the solenoid of the typewriter. A Uher duoplot system was used to start a millisecond clock at the onset of each auditory and visual word.

In Experiment I, *Ss* were required to decide whether or not each word had occurred before in the list. They responded by saying *yes* or *no* on each trial in the continuous list. The RT was measured from the onset of each word, and the clock

was stopped by S's vocal responses amplified from a microphone signal.

In Experiment II, *Ss* were required to say *same*, *different*, or *no*. If a word had occurred before in the list, *Ss* said *same* or *different* contingent on their decision as to whether or not the word had been presented in the same physical form on the 2 occasions. The *Ss* said *no* if the word had not appeared in the list before. The RT was not measured in this experiment.

RESULTS

Recognition accuracy. The recognition data for the inter- and intramodality combinations are presented in Table 1. As there was no significant effect attributable to the task variable, the accuracy data are shown pooled across *Ss* in the RT and modality judgment experiments. Each value is therefore based on 600 observations representing 15 observations per *S* for the 40 *Ss* in the RT and modality judgment experiments.

The auditory comparison reveals an advantage for AA relative to VA for all 6 retention intervals. An analysis of variance (ANOVA) (Task × Similarity × Lag) showed that the main effects of similarity, F (1, 38) = 17.3, $p < .001$, and lag, F (5, 190) = 34.7, $p < .001$, were significant. There were no other significant effects (at the .05 level). Thus, the tabular and statistical evidence are consistent with the conclusion that the AA advantage does not diminish over the sampled range of retention intervals.

The visual comparison shows a consistent advantage for VV at the 4 longest retention intervals and some inconsistency at the 2 shortest intervals. An ANOVA (Task × Similarity × Lag) showed that the main effects of similarity, F (1, 38) = 4.8, $p < .05$, and lag, F (5, 190) = 30.5, $p < .001$, were significant. The Similarity × Lag interaction was also significant, F (5, 190) = 8.7, $p < .001$, indicating that the VV advantage increases significantly at the longer retention intervals. There were no other significant effects.

An overall ANOVA (Task × *Ss* × Test Modality × Similarity × Lag) was completed in order to test the hypothesis that the advantage found for the intramodality conditions followed different retention functions in audition and vision. The critical Test Modality × Similarity × Lag interaction, F (5, 190) = 6.7, $p < .001$, showed that this was the case.

The overall performance level was similar in the 2 experiments. The mean detection and false alarm rates were 91.0% and 10.5% in Experiment I and 92.9% and 6.6% in Experiment II. Thus, the data show that recognition performance was not impaired by the additional input and output demands associated with the judgment task in Experiment II.

Recognition reaction time. The results presented in Table 2 show recognition RT for correct recognition decisions for the 4 modality combinations. The values are based on not less than 233 observations per point.

For the auditory comparison, the recognition RT data show a pattern similar to that for recognition accuracy. The auditory-auditory combination enjoys a consistent advantage at most retention inter-

TABLE 1

PROPORTION OF WORDS CORRECTLY RECOGNIZED AS A FUNCTION OF MODALITY COMBINATION AND LAG

Modality combination	Lag						
	[illegible]	[illegible]	4	8	16	[illegible]	[illegible]
AA	98.8	99.0	[illegible]	[illegible]	89.1	[illegible]	[illegible]
VA	97.8	[illegible]	96.3	[illegible]	88.3	[illegible]	96.9
VV	98.8	95.3	[illegible]	91.0	[illegible]	[illegible]	[illegible]
AV	98.6	98.8	94.3	[illegible]	85.9	[illegible]	90.9

[illegible]

TABLE 2

RECOGNITION REACTION TIME (IN MSEC.) FOR WORDS CORRECTLY RECOGNIZED AS A FUNCTION OF MODALITY COMBINATION AND LAG

Modality combination	Lag						
	1	2	4	8	16	32	M
AA	786	829	824	918	926	1017	883
VA	788	898	932	968	971	965	920
VV	713	764	797	763	774	805	769
AV	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Note. Abbreviations: AA = auditory auditory, VA = visual-auditory, VV = visual visual, and AV = auditory visual.

vals, however, unlike the accuracy data, the AA advantage is reversed at the longest retention interval and substantially reduced in magnitude at the shortest interval. An ANOVA (Similarity × Lag) showed that the AA advantage, $F(1, 19) = 9.4, p < .01$, lag effect, $F(5, 95) = 31.4, p < .001$, and the Similarity × Lag interaction, $F(5, 95) = 5.1, p < .001$, were all statistically significant.

The visual comparison supported the accuracy data with a clear VV advantage at the longest retention intervals and a lesser advantage at the shorter intervals. An ANOVA showed that the effects of similarity, $F(1, 19) = 14.0, p < .01$, lag, $F(5, 95) = 8.8, p < .001$, and Similarity × Lag interaction, $F(5, 95) = 2.5, p < .05$, were significant.

To summarize, the evidence shows that recognition performance is sensitive to the modality manipulation. In the case of auditory persistence, the tabular data indicate an AA advantage relative to VA at the 5 shortest retention intervals for both accuracy and recognition RT. The recognition RT data indicated that auditory persistence might be limited to a period of less than 2 min. or 31 intervening items, but this may be due to technique-specific or sampling factors. The data clearly supported the conclusion that there is an overall advantage for VV relative to AV and suggested that, if anything, this may be enhanced at long retention intervals.

Knowledge of modality of presentation. The data presented in Table 3 show the proportion of correct judgments for each combination of the experimental conditions. The data are necessarily conditionalized on correct recognition. The results of Experiment II show an advantage for the intramodality conditions, particularly at long retention intervals. An ANOVA (Test Modality × Similarity × Lag) showed that the main effects of Similarity, $F(1, 19) = 11.4, p < .01$, and lag, $F(5, 95) = 27.9, p < .001$, were significant. The Test Modality × Similarity × Lag interaction, $F(5, 95) = 5.5, p < .001$, was also significant.

There are 2 possible sources of artifact in the modality judgment data; first, *S*s could improve their performance by giving more *no* responses when uncertain about modality of first presentation and, secondly, *S*s might give more *same* responses to items in one or both modalities. The finding that accuracy was greater in the modality report group appears to preclude the general criterion effect.

The response bias problem is more complicated. The false alarm data presented in Table 3 indicate a tendency to give more *same* than *different* responses, particularly for visual items. A signal detection analysis was implemented to give an unbiased estimate of modality knowledge. In this analysis, the d' statistic was calculated separately for each condition for each *S*. The hit rate was the number of correct modality judgments expressed as a proportion of the number of correct detections within each condition. The false alarm rates were calculated from the allocation of *same* and *different* judgments to new items. For this purpose it was assumed that the population of new items was

TABLE 3

Proportion of Correct Modality Judgments as a Function of Modality Combination and Lag

Modality combination	Lag: 1	2	4	8	16	32	*M*	False alarms[a]
AA	99.7	99.7	97.9	91.7	89.2	80.0	93.0	3.4
VA	98.3	90.7	96.3	89.2	86.0	77.7	89.7	3.1
VV	99.0	92.3	94.7	89.9	87.3	86.0	91.5	3.9
AV	98.3	99.7	94.6	85.3	84.9	76.7	89.9	2.7

Note. Abbreviations: AA = auditory–auditory, VA = visual–auditory, VV = visual–visual, AV = auditory–visual.
[a] For example, 3.4 and 3.1 indicate the proportion of *same* and *different* responses, respectively, given to new auditory items.

divided equally into *same* and *different* events. Thus, for the VV condition, the false alarm rate was the number of *same* judgments given to new visual items. The same procedure was adopted for *different* judgments. The d' values were 4.04, 3.94, 3.96, and 3.96 for the AA, VA, VV, and AV modality combinations, respectively. An ANOVA completed on the d' values showed that the main effect of similarity was not significant ($F < 1$), but the Test Modality × Similarity × Lag interaction, F (5, 95) = 4.1, $p < .01$, remained significant. Thus, the effect of this calculation was to eliminate the overall advantage associated with the intramodality conditions.

Discussion

The results show that information about modality of presentation remains in memory for a period that must be considered in minutes rather than seconds. This is evident from both the direct evidence that *S*s can give reliable judgments about modality of presentation over a 2-min. period, and from the indirect evidence that recognition performance is facilitated when a word is presented in the same modality on the presentation and test trials.

Detailed analysis of the recognition data indicated some asymmetry in the retention of auditory and visual material. Auditory persistence was at a maximum over short retention intervals of 16 items (64 sec.) or less, but the data did not consistently support the view that auditory persistence was of limited duration. Visual persistence, on the other hand, appeared to be masked at short retention intervals but was present under all conditions at retention intervals beyond 4 intervening items or 16 sec. Unlike recognition, however, the modality judgment data showed that when response biases were controlled there was no evidence of either (*a*) asymmetric retention of auditory and visual information or (*b*) an advantage attributable to the presentation of a word in the same modality on the presentation and test trials.

The most important theoretical questions concern the way in which modality information is encoded and stored in memory and how it is retrieved and used in the recognition and modality report situations. There appear to be a number of possible explanations of the data, ranging, at one extreme, from the view that they are influenced by an extended form of sensory memory, to the view that recognition facilitation as well as modality knowledge is based on abstract information stored in long-term memory.

As argued by Serfass and Dong (1972), the fact that *S*s can remember input modality is not readily explained by reference to information-processing models such as those advanced by Atkinson and Shiffrin (1968), Morton (1969), or Sperling (1967). Certainly, it appears true that these models have neglected the processes involved in storing attributional information about an item in favor of an analysis of the operations involved in the identification and retention of the item per se. However, it may be more appropriate to address modality report data to the question of memory for attributes raised by Underwood (1969). In this context the modality report data may not be critical for any one model of attribute retention.

Hintzman, Block, and Inskeep (1972) have considered 2 explanations of attribute report data. The first of these is that an attribute is encoded as an abstract proposition, associated with the meaning of the word, and bearing only a referential relationship to the stimulus conditions. The second is that attribute information is retained as a literal copy of the perceptual experience. The present data bear on these issues at a number of points. First, as Hintzman et al. have argued, if the modality information is stored as an abstract attribute it should make no difference to recognition whether or not an item is re-presented in the same modality. In this respect the present results clearly support the literal or "representational" position. Furthermore, the asymmetric relationship between auditory and visual persistence in recognition may be adduced as evidence against the view that modality information is encoded only in an abstract form. It is possible, however, that attribute information is encoded and used in different ways for recognition and attribute report. Thus, modality report could be based on abstract information even though recognition facilitation is determined in some other way.

The abstract explanation of the report data received some support from the finding that modality knowledge is generally insensitive to the similarity manipulation. Unlike recognition, performance on the modality report task is not facilitated in the same modality condition. On the other hand, there is evidence that performance on the between-modalities

attribute is superior to performance on within-modalities attributes. This trend was reported by Hintzman et al. (1972), and is apparent in a comparison between the results of the present experiment and the results of (*a*) an analogous visual experiment using 2 type fonts (Kirsner, 1973), and (*b*) an auditory experiment using 2 voices (Craik & Kirsner).[3] Thus, if attribute information is stored only in an abstract form dependent on the verbal unit in memory, then it seems necessary to conclude that the organization of the structure surrounding the word unit is influenced by perceptual as well as linguistic factors.

In summary, it has been argued that (*a*) modality information is stored in a representation form in memory, (*b*) this representational information is used directly in recognition, and (*c*) modality judgments also use information which may be stored in a representational form, but that access to this information is achieved via the word unit. These conclusions are of course tentative, and an accurate description of the relationship between representational and linguistic systems in memory depends on further research.

[3] F. I. M. Craik and K. Kirsner. The effect of speaker's voice on recognition. In preparation.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. *Advances in research and theory.* New York: Academic Press, 1968.

Burrows, D. Modality effects in retrieval of information from short-term memory. *Perception & Psychophysics*, 1972, **11**, 365–372.

Hintzman, D. L., Block, R. A., & Inskeep, N. R. Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 741–749.

Kirsner, K. An analysis of the visual component in recognition memory for verbal stimuli. *Memory & Cognition*, 1973, **1**, 449–453.

Kirsner, K., & Craik, F. I. M. Naming and decision processes in short-term recognition memory. *Journal of Experimental Psychology*, 1971, **88**, 149–157.

Kroll, N. E. A., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. Short-term memory while shadowing: Recall of visually and aurally presented letters. *Journal of Experimental Psychology*, 1970, **85**, 220–224.

Mann, J. E. Modality effects in free recall ordered by presentation mode. Unpublished bachelor's thesis, University of Toronto, 1970.

Morton, J. A functional model for memory. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1969.

Murdock, B. B., Jr. Auditory and visual stores in short-term memory. *Acta Psychologica*, 1967, **27**, 316–324.

Murdock, B. B., Jr., & Walker, K. D. Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 665–676.

Parkinson, S. R. Short-term memory while shadowing: Multiple-item recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, 1972, **92**, 256–265.

Parks, T. E., Kroll, N. E. A., Salzberg, P. M., & Parkinson, S. R. The persistence of visual memory as indicated by decision time in a matching task. *Journal of Experimental Psychology*, 1972, **92**, 437–439.

Serfass, C. P., & Dong, T. Retention of item and modality information in serial word lists. Paper presented at the meeting of the Midwestern Psychological Association, Cleveland, Ohio, May 1972.

Shepard, R. N., & Teghtsoonian, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, **62**, 302–309.

Sperling, G. Successive approximations to a model for short-term memory. *Acta Psychologica*, 1967, **27**, 285–292.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

Warrington, E. K., & Shallice, T. The selective impairment of auditory verbal short-term memory. *Brain*, 1969, **92**, 885–896.

(Received May 7, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 585–594

THE EFFECT OF LEVEL OF DEPTH PROCESSING AND DEGREE OF INFORMATIONAL DISCREPANCY ON ADAPTATION TO UNIOCULAR IMAGE MAGNIFICATION [1]

WILLIAM EPSTEIN [2] AND CYNTHIA L. MORGAN-PAAP

University of Wisconsin—Madison

The *S*s adapted to uniocular image magnification which transformed the preexperimental relationship between objective depth slant and retinal disparity. The adaptation treatment consisted of continuous inspection of a single luminous figure representing for different *S*s three different degrees of perspectival slant. Two levels of depth processing were contrasted: *depth registration*—an activity which diverted *S* from active utilization of depth information—and *depth processing*—an activity that required frequent utilization of depth information. The magnitude of adaptation was found to increase as the discrepancy between perspectival and stereoscopic slant increased and to be greater following depth processing than following depth registration. The results were discussed in the context of an account of adaptation which stresses resolution of informational discrepancy. Other findings in the literature were reinterpreted in the light of the current experiments.

In earlier experiments (Epstein, 1971, 1972a, 1972b; Epstein & Daviess, 1972; Epstein & Morgan, 1970) we showed that the relationship between retinal disparity and perceived stereoscopic depth is modified by a period of exposure to uniocular horizontal image magnification produced by a meridional size lens (MSL). The *S*s, who engaged in binocular visual exploration of a typical indoor environment, while wearing a vertically oriented MSL before one eye, showed subsequent adaptive shifts in depth judgments based on binocular disparity. The shifts were interpreted as evidence that binocular disparity had been recalibrated as a consequence of the pairing of disparity with the nontransformed discrepant information provided by exposure to the normal environment.

The present experiments had three objectives. First, we wished to test a procedure for bringing the adaptation treatment under experimental control. Although the early contemporary research (reviewed by Epstein, 1967, Chap. 9) employed controlled laboratory tests to produce adaptation, e.g., target-marking, much recent work, including our previous work on uniocular image magnification, has favored the less controlled expedient of giving *S* a period of exposure to the normal extralaboratory environment. The advantage of this latter treatment is that it combines all of the nontransformed informational inputs from the environment in joint opposition to the optically transformed input, thus providing optimal opportunity for adaptation. The disadvantages are all those associated with lack of control. Therefore, in Experiment II we set about to demonstrate adaptation to uniocular image magnification with a strictly controlled treatment.

Our second objective was to examine the relationship between the degree of informational discrepancy and the magnitude of the adaptive shift. In the present experiments the discrepancy was between perspective slant and stereoscopic slant. We chose perspective as the source of discrepant input because there is evidence that in the absence of other information, perspectival cues are effective determinants of perceived slant in depth (e.g., Clark, Smith, & Rabe, 1955; Freeman, 1966), and because

[1] This investigation was supported by U.S. Public Health Service Grant MH-16390. The authors wish to thank Krishan D. Broota, S. M. Ebenholtz, and Kenneth Paap for help with various aspects of this investigation.

[2] Requests for reprints should be sent to William Epstein, Department of Psychology, University of Wisconsin—Madison, Psychology Building, Madison, Wisconsin 53706.

there is evidence that in a cue-conflict situation perspective can modify the effect of binocular disparity (Gillam, 1968). Despite the aforementioned evidence, Experiment I was designed to assess *S*'s depth judgments in a cue-conflict situation involving disparity and perspective. The choice of perspective is also recommended upon considerations of experimental convenience. It is a simple matter to vary perspectival slant and thus vary the degree of informational discrepancy. In Experiment II, stereoscopic slant, determined by the MSL, was kept constant, while perspectival slant, determined by the figural properties of the target that was inspected during the adaptation period, was varied. The adaptation treatment consisted of exposure to discrepancy between perspectival slant and stereoscopic slant, and the measure of adaptation was the shift in binocular, no-lens, judgment of the apparent frontal parallel plane of a single luminous rotatable test line.

A third objective was to examine the influence of level of processing on adaptation. It now is commonplace that adaptive shifts are greater following self-generated, self-controlled movement than following passive, other-controlled movement. Although a variety of interpretations of this result have been advanced, there is a relatively straightforward account which rarely has been made explicit. On the view that adaptation represents a response to informational discrepancy, it is plausible that the degree of adaptation would depend on the degree of processing of the inputs that are potentially discrepant. If processing of the input has no behavioral significance, because behavior is guided externally, as in the case of passive movement, then processing may be aborted and the informational discrepancy may not be sufficient to motivate adaptation. Accordingly, adaptation following passive activity will be less than adaptation following self-generated activity.

In the foregoing account, the important difference is not between active and passive movement, but between levels of processing; simple registration of input versus processing of the input in the service of a relevant perceptual discrimination. Therefore, in Experiment II we examined the effects of two types of exposure tasks on adaptation to uniocular image magnification. In both tasks the displays, conditions of viewing, and duration of visual exposure were identical. In the *depth registration* condition *S*'s attention was drawn to a task that did not require utilization of depth information; in the *depth-processing* condition *S*'s attention was drawn to a task that required processing of depth information. Our expectation was that *depth processing* would lead to greater adaptation.

Experiment I

Prior to the investigation of our principal questions it was necessary to assess the responses of the perceptual system to the discrepancy between perspectival and stereoscopic slant. What is the apparent slant in depth of a target whose retinal perspectival correlates and retinal-disparity correlates are not in agreement? To answer this question, judgments were secured of the apparent slant of frontal parallel projections of a rectangle rotated in depth about its vertical axis, presented under three conditions of viewing: monocular viewing (M), binocular untransformed viewing (B), and binocular viewing with MSL before one eye (MSL). On some trials in the MSL condition, the direction of the theoretical rotation of the frontal plane induced by the MSL, i.e., the direction of stereoscopic slant expected in the absence of conflicting input, was consistent with the direction of perspectival slant; on other trials, the directions of perspectival and stereoscopic slant did not agree. With these variations in mind, the specific questions to be resolved in Experiment I were the following: (*a*) In the absence of discrepant input (Condition M), will the perspectival cue determine apparent slant-in-depth reliably? (*b*) When normal access to binocular disparity is available (Condition B), will the effect of perspective be attenuated? (*c*) When the discrepancy between stereoscopic slant and perspectival slant is accentuated (Condition MSL), is

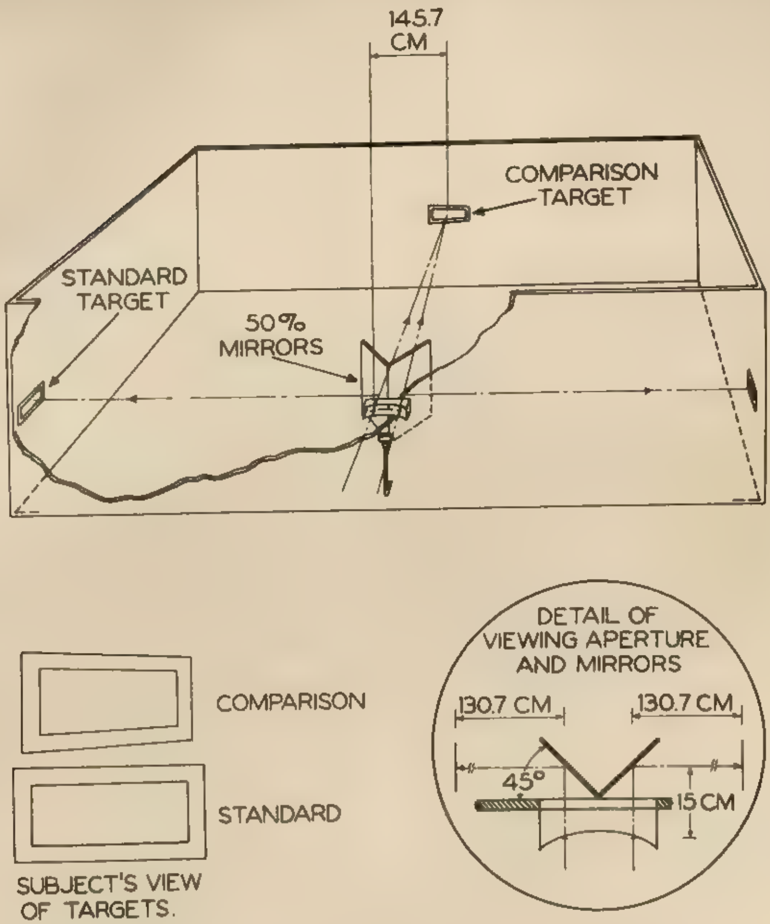


FIGURE 1. Arrangement for presenting stimuli. (Explanation is in text.)

the result a perceptual compromise or does one or the other of the two inputs dominate completely?

Method

Apparatus. Figure 1 shows the arrangement for presenting the stimuli. The *S* always viewed two targets simultaneously, one above the other: a monocularly viewed standard presented to the left eye and a comparison which was presented for either monocular or binocular inspection as the experiment required. The comparison target was 145.7 cm. directly in front of *S*. The 50% mirrors just inside the viewing aperture at 45° angles allowed the monocularly seen standard to appear at the same distance and directly below the comparison. The standard and comparison targets were equivalent in subjective brightness. Even though the monocular target was viewed only with the left eye, a mirror was placed in front of both eyes so that both eyes would receive similar amounts of light. The *Ss* viewed the targets through an aperture 13.75 × 6.25 cm. while supported by a chin rest.

Stimuli. The targets were photographic negatives of a 20.4 × 10.2 × 1.9 cm. rectangle, photographed at slants of −60° to +60° in 10° steps. A 20.4 × 10.2 cm. rectangle subtends a visual angle of 8° × 4° at the viewing distance used. The targets were placed in slots in front of frosted glass and were back-illuminated. They appeared to *S* to be luminous white against a totally black background. In addition, there was a horizontal line target, 20.4 × 6 cm. The line was cut into a black surround and when back-illuminated like the other targets, it too appeared to be a luminous white against a black background.

Lens. A 6% meridional size lens and two plain glasses ground by the American Optical Company were used.

Design. Twelve presentation orders were generated from a Latin square. Each of the 12 slants of the comparison (±10° to ±60°) appeared in each serial position only once and each slant was preceded by and followed by every other slant only

once. The zero-slant rectangle was always the first and the last comparison target in each viewing condition. It was the only target presented twice. The standard was always a rectangle, a zero-slant target. Slant estimates were secured for each comparison relative to the standard. Every *S* judged all comparison slants under each of the three viewing conditions in succession: M, B, and MSL. The order of viewing conditions was determined randomly for each *S*.

Procedure. The *S* was seated with head in a chin rest, at the viewing aperture. The viewing aperture was parallel with the forehead and eye level was at the height of the midpoint of the aperture. First the effectiveness of the MSL was determined. This was accomplished by presenting the horizontal line for binocular inspection in the place where the comparison targets would later be seen. The monocular standard was also in view and *S* was instructed to provide an estimate of the slant of the line relative to the standard using a reproduction procedure:

> You will see a rectangle and a horizontal line, one above the other. The rectangle will be of zero slant, that is, parallel to your forehead. Your task will be to reproduce the apparent visual slant of the line relative to the rectangle; that is, to draw an angle which corresponds to the apparent angular separation of the rectangle and the line. I am only interested in how slanted the line appears to be, and you need not worry about where it actually is.

The *S* was given a clipboard with sheets of paper attached to it. Each sheet showed a circle 11.75 cm. in diameter with a small dot marking its center and a line drawn through its center bisecting the circle. Demonstrations of drawings of the line to reproduce the angular rotation of the horizontal line were provided.

Throughout the entire experiment *S* wore an American Optical trial frame. For the screening test a plain glass was inserted in one frame while the other had the 6% MSL. The MSL eye was the one subsequently to be used in the MSL condition for that *S*. The *S* observed the line and rectangle for 45 sec. before responding. The 45-sec. viewing time was chosen because the stereoscopic effect of the MSL is not immediate (Seagrim, 1969). The lights were then turned on, *S* removed the glasses, and then drew the apparent angular rotation of the horizontal line on the top sheet of paper.

Then *S* was exposed to the three viewing conditions and the succession of 14 targets in each condition. In Condition M, *S* wore a plain glass before the left eye and an eye patch over the right eye, in Condition B the trial frame had two plain glasses (PG), and in Condition MSL the MSL was before one eye and a PG was before the other. Half of the *S*s wore the lens before the right eye and half before the left eye. The following instructions were read:

> You will see two rectangles, one above the other. The top rectangle will be changed in slant from trial to trial at random but the bottom will always be a rectangle of zero slant, that is parallel to your forehead. Your task will be to reproduce the apparent visual slant of the top rectangle relative to the bottom rectangle, that is, to draw an angle which corresponds to the apparent angular separation of the two rectangles. I am only interested in how slanted the top rectangle appears to be, and you need not worry about where it actually is.

The room lights were turned on while *S* reproduced the angular rotation of the comparison and changed the targets. After each reproduction *S* moved the response sheet so that preceding reproductions were not open to inspection by *S*.

Subjects. The 24 *S*s were University of Wisconsin undergraduates. Persons who normally wore framed glasses were not selected but wearers of contact lenses were accepted. To secure these 24 *S*s, 34 were tested. Ten were rejected because they did not exhibit an MSL effect when viewing the horizontal line.

Results and Discussion

On the screening test, slant estimates of the horizontal line were 30.83° ($SD = 9.09°$) for the 12 *S*s who wore the MSL before the right eye, and also 30.83° ($SD = 8.16°$) for the 12 *S*s who wore the MSL before the left eye. The theoretical geometric effect (Ogle, 1964) for the 6% MSL at an observation distance of 145.7 cm. is 51°5′. Therefore, in estimating the slant of an objectively frontal parallel horizontal line in the absence of veridical slant cues *S* should estimate it as being slanted 51°5′ with the side in front of the lens farther away. Although no *S* exhibited the full effect, all 24 *S*s did exhibit large effects in the predicted direction. The principal data are summarized in Table 1. Analysis of variance showed that neither MSL eye (left or right) nor condition of viewing (M, B, or MSL) were significant sources of variance. However, the effect of degree of perspectival slant was highly significant, F (6, 132) = 769.13, $p < .001$. Table 1 shows that each interval of increased perspectival slant was accompanied by a corresponding increase in estimated slant. Also significant was direction of perspectival slant, F (1, 22) = 6.05, $.025 > p > .01$, although this effect was much smaller than the effect of magnitude of slant. The targets designed to appear with the left side farther away were

...erally estimated at larger slants than ...rgets of the same magnitude slanted with ...right side farther, although the differ-...s between a left- and right-slanted ...get of the same magnitude was small. Also significant was the Direction × Magnitude of Slant interaction, F (6, 132) = ...69, $.05 > p > .01$, and the MSL Eye × ...ection × Magnitude of Slant inter-...on, F (6, 132) = 2.45, $.025 > p > .01$.

The three questions that Experiment I ... designed to answer have been answered ...quivocally. Under the experimental ...ditions, perspectival input had significant and systematic effects on perceived slant depth. The effectiveness of the perspectival input was in no way modified by conflicting stereoscopic slant depth. Perspectival slant was clearly dominant.

Experiment II

The results of Experiment I provided a ...asis for inferring the degree of informa-...ional discrepancy that will be present during an adaptation period in which Ss are exposed to a constant stereoscopic slant coupled with varying degrees of perspec-...ival slant. Consider the case of an S who spends 10 min. in binocular inspection of one of the comparison stimuli of Experiment I while wearing an MSL before one eye. Suppose that the geometric effect of the MSL is a 60° rotation. If the inspection figure is a rectangle (zero slant), the discrepancy between perspectival and stereoscopic slant will be 60°, but if without changing the MSL, the inspection figure is changed to a 60° projection in the same direction as the MSL-induced slant, the discrepancy will be reduced to zero. And by choosing an intermediate projection, e.g., 30°, an intermediate degree of discrepancy is introduced. This plan was followed in Experiment II. Three groups of Ss were tested for adaptation to uniocular image magnification following a period of inspection of one of the perspectival projections used in Experiment I. It was expected that the magnitude of the adaptive shift would be larger the greater the degree of discrepancy, despite the fact that disparity was transformed to the same degree

TABLE 1

Mean Estimates of Slant of Comparison Targets under the Three Viewing Conditions

Comparison	Monocular	Binocular: With MSL[a]	Binocular: Without MSL	Grand M
0°	0	0	0	0
10°	9.50	9.31	9.08	9.30
20°	20.40	19.25	19.00	19.55
30°	29.38	28.75	28 75	29.00
40°	37.75	37.50	37.50	37.50
50°	46.70	47.00	46.50	46.75
60°	56.63	57.50	57.04	57.00

[a] MSL = meridional size lens.

for all Ss. Each S adapted twice, once in the *depth-registration* condition and once in the *depth-processing* condition.

Method

Apparatus. The apparatus used in Experiment I was employed in Experiment II with a few alterations. The viewing distance to the comparison target was increased from 145.7 cm. to 154.48 cm. The slot where the standard target was shown in Experiment I was covered and not used in Experiment II. Only the comparison target area was used to present the adaptation figures. At a distance of 101 cm. to the left of the chin rest used to look at the targets in Experiment I, a second chin-rest and head-clamp arrangement was set up on the same level as the first chin rest. At a distance of 154.48 cm. a rotatable electroluminescent line, 29 × .5 cm., was located. The height of the line was at S's eye level. The line could be rotated in the horizontal plane on a vertical axis about a pivot point which remained fixed directly in front of S. The position of the line could be read to an accuracy of .5° of rotation. A reading of 180° signified that the line was perpendicular to S's line of sight, that is, in the frontoparallel position; a reading over 180° indicated that the left side was rotated back; a reading under 180° indicated that the right side was rotated back. The chin rest and head clamp inhibited large head movements by S. In order to increase the range of discrepancies that could be studied in our experimental arrangement, an 8% MSL was used in Experiment II. At the viewing distance of 154.58 cm. the geometric effect of the MSL is an apparent rotation of 60°. The trial frame used in Experiment I was used again in Experiment II.

Stimuli. The inspection figures presented to S during the adaptation period were the 0°, 30°, and 60° figures previously used in Experiment I.

Design. Three independent groups of 20 Ss were tested. Each S was exposed to only a single adaptation figure, but each S participated in both adaptation activities. Within each group, half of the Ss

had the depth-registration activity on the first day and the depth-processing activity on the second day, while the other *S*s had the two adaptation activities in reverse order. The two adaptation sessions were separated by 48 hr. Half of the *S*s in each group wore the MSL before the right eye during the adaptation period and half wore the MSL before the left eye. The adaptation figures and the MSL eye were always paired together in such a way that perspectival slant and stereoscopic slant were in the same direction.

Procedure. All *S*s followed the identical preadaptation procedure. First, *S* made eight judgments of the frontoparallel position of the electroluminescent line. The *S* was seated at the chinrest and head-clamp arrangement and viewed the electroluminescent line binocularly in total darkness. The first four judgments were made wearing the trial frame with PG in both frames. The second four judgments were made wearing the trial frame with a PG in one frame and the 8% MSL in the other. Two of the settings of each set of four started with the line at 155° and two with the line at 205°.

The *S* instructed *E* to adjust the position of the line until it appeared parallel with her forehead. Head movements were explicitly discouraged. In order to reduce the tendency to make judgments based on line length rather than slant depth, *S*s were told to disregard the length of the line and judge exclusively on the basis of the line's orientation relative to her forehead. This instruction was necessary since, with a predicted geometric effect of 60°, apparent line length could become a confounding variable.

Following the preadaptation tests, the glasses were removed and then *S* moved to the other chin rest. The *S*s in the *depth-registration* condition heard the following instructions:

> This stage of the experiment is concerned with signal detection. We wish to learn whether the ability of an observer to detect a faint signal varies over a ten-minute period of observation of a simple unstimulating visual field. When you look into this box you will see a luminous four-sided figure in an otherwise dark background. At infrequent and irregular intervals a faint light will be flashed inside the figure. The light will appear in one of the four corners. The corners at which the light will appear have been selected randomly, so it will not be possible for you to predict where the signal will appear. When you have detected a signal press one of the four buttons to show its location in the figure.

The *S* was shown how to use the response box buttons. The glasses were restored with the MSL still in place and then she observed the figure and responded for 10 min. The *S*s in the *depth-processing* condition heard the following instructions:

> This stage of the experiment is concerned with another form of depth judgment. We wish to learn whether the depth at which objects app[ear] varies over a ten-minute period of continuo[us] observation. When you look into this box [you] will see a luminous four-sided figure in an ot[her]wise dark background. At infrequent [and] irregular intervals a faint light will be flash[ed] inside the figure. The light will appear in [one] of the four corners. The corners at which th[e] light will appear have been selected randomly so it will not be possible for you to pred[ict] where the signal will appear. Each time a ligh[t] flashes I want you to answer two ques[tions] about the corner in which the light appear[s]: (*a*) Does the corner, that is, the corner wh[ere] the light appeared, appear nearer, farther [or] equal in distance from you as the corner to [the] right or left? (*b*) Does the corner app[ear] nearer, farther, or equal in distance from yo[u] as the corner below or above it?

An example was presented and when *S* understo[od] the procedure, the glasses were restored with th[e] MSL in place and she observed the figure and re[s]ponded for 10 min.

The light flashes were brief and appeared [at] random times at intervals ranging from 30 sec. [to] 90 sec. in randomly ordered corners. After the 10-min. adaptation activity, *S* kept her eyes closed while *E* changed the lenses and while she returned to the other chin rest. All *S*s made two PG postadaptation judgments of the frontoparallel position with the electroluminescent line. Forty-eight hou[rs] later, the procedure was repeated with the second adaptation activity.

Subjects. Sixty undergraduates from the sa[me] population that provided *S*s for Experiment I serv[ed] in Experiment II. Twenty *S*s were assigned to each of the three adaptation figures. Since the principal objective of this investigation was to examine the recalibration of binocular disparity as a stimulus for perceived depth we could use as *S*s only observers who gave evidence of effective use of disparity prior to the introduction of the experimental treatment. To secure 60 *S*s who satisfied this requirement it was necessary to test a total of 97 *S*s. Twenty-four were rejected because the mean of the 4 frontoparallel PG settings deviated from the true frontoparallel by more than ±10°; 11 were rejected because of failure to exhibit an MSL effect of at least 20°; and 2 were dropped due to a failure to return for the second day of testing. With the exception of the latter 2 *S*s, all rejected *S*s were terminated before the initiation of the adaptation period on the first day. None of the retained *S*s wore framed corrective glasses but several were regular users of contact lenses and these *S*s continued to wear the contact lenses throughout the experiment.

Results

Preexposure plain glass judgments. The results are summarized in Table 2. The

TABLE 2

Mean Preadaptation Setting of Frontal Parallel, Mean Shift, and Percept Adaptation: Depth Registration First; Depth Processing First

Level of processing	0° adaptation figure			30° adaptation figure			60° adaptation figure			Average % adaptation
	Pre M	Shift M	% adaptation	Pre M	Shift M	% adaptation	Pre M	Shift M	% adaptation	
Depth registration first										
Depth registration	181.02	19.48	56.86	179.78	14.40	46.58	180.38	10.90	34.48	45.97
Depth processing	180.72	26.22	74.24	180.85	22.65	74.20	180.48	16.88	53.50	67.31
Depth processing first										
Depth registration	179.58	16.50	49.80	180.60	11.40	35.21	180.55	6.90	21.31	35.44
Depth processing	180.52	24.52	69.84	179.92	15.52	49.05	180.65	9.62	30.43	49.77

mean preexposure frontoparallel judgments, based on the four settings, show that *S*s were highly accurate. A between-*S*s (adaptation figure and MSL eye), within-*S*s (adaptation activity) analysis of variance of the preexposure settings showed no significant main effects or interactions.

Preexposure meridional size lens judgments. The MSL produces an apparent rotation of the frontal plane, so that when the electroluminescent line is objectively in the frontoparallel position, the side on the same side as the lens appears farther back. Therefore to adjust the line to the apparent frontoparallel plane the right side of the line must be moved forward when the right eye is exposed to the lens and the left side of the line must be moved forward when the left eye is exposed to the lens. As noted earlier, the predicted geometric effect is 60°.

All *S*s made adjustments in the predicted direction. When the line was judged to be in the frontoparallel position, the mean MSL settings differed from the mean PG settings (the MSL effect) by 34.38°, 31.47°, and 32.93° for the groups assigned to the 0°, 30°, and 60° adaptation figures, respectively. A between-*S*s (adaptation figure and MSL eye), within-*S*s (adaptation activity) analysis of variance was computed to determine whether the groups differed before the start of the different adaptation treatments. There were no significant main effects. There was one significant interaction: Adaptation Figures × Adaptation Activity, $F\ (2, 54) = 3.42$, $p < .05$.

Postadaptation judgments. Table 2 shows the mean postadaptation plain glass shift (difference between the pre- and postadaptation judgments) and the percent adaptation for the three adaptation figures and the two adaptation activities. Adaptation is evident if at the removal of the MSL *S* judges the line to be frontoparallel when the line is set to produce the same disparity that was produced by the MSL. Therefore, the effect will be a deviation of judgments of frontoparallel from the preadaptation frontoparallel, with the side of the line on the same side occupied formerly by the MSL set farther back.

All *S*s showed aftereffects in the predicted direction with mean shifts ranging from 6.90° to 26.22°. The degree of shift depended on the inspection figure and the type of adaptation activity. The effect of these variables is readily discerned in Figure 2 which presents percent adaptation averaged over the two orders in which the two adaptation activities were experienced. Figure 2 shows that the magnitude of the adaptive shift decreased as the perspectival slant of the inspection figure increased and that for all inspection figures the adaptive shift was greater following the depth-processing activity compared to the depth-

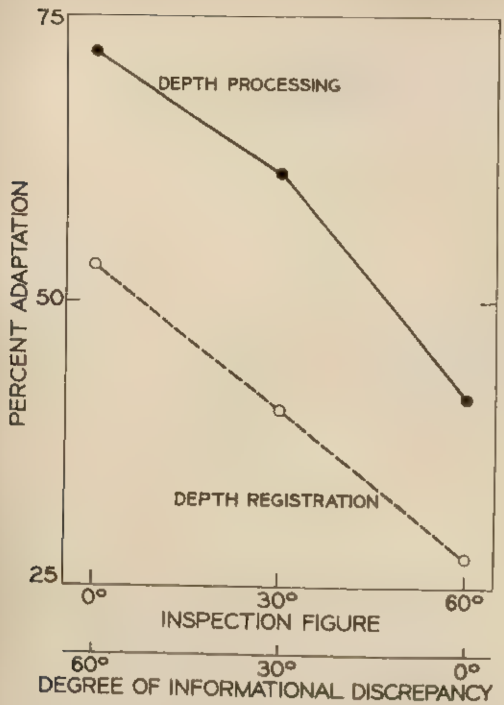


FIGURE 2. Percent adaptation for three inspection figures (degree of discrepancy) and two types of adaptation activity averaged for the two orders.

registration activity. The average percent adaptation was 63%, 51%, and 35% for the 0°, 30°, and 60° inspection figures, respectively. The average percent adaptation was 64% and 45% for the depth-processing and depth-registration activities, respectively. A between-*S*s (adaptation figure, MSL eye, and order of adaptation activity), within-*S*s (adaptation activity) analysis of variance was performed on the percent adaptation data. The main effects of adaptation figure, F (2, 48) = 19.88, $p < .001$, MSL eye, F (1, 48) = 4.76, $p < .05$, adaptation activity order, F (1, 48) = 15.10, $p < .001$, and adaptation activity, F (1, 48) = 90.48, $p < .001$, were significant. There were no significant interactions. The effect of MSL eye was due to the slightly greater adaptation for *S*s who wore the MSL before the left eye. The order of activity effect was due to the fact that the passive–active order resulted in greater adaptation than the reverse order.

Discussion

Experiment II showed that for a con transformation (enhancement) of bin disparity the magnitude of the adaptive varied as a consequence of variations of spectival slant of the adaptation figure. Th are no strictly analogous observations in literature. However, our findings can related to the results of experiments that h varied the degree of optical transformat e.g., degree of optical tilt (Ebenholtz, 1966 degree of sideways displacement (Efstath 1969). In these cases, the adaptive shift been found to increase as the magnitude of t optical transformation has increased. Th present findings strongly suggest that the essen tial variation, common to both procedures, i a variation in the degree of informational discrepancy. The magnitude of discrepancy can be varied by manipulating the degree of optical transformation or by varying the informatio correlated with the nontransformed input The effect is comparable. Nevertheless, ou procedure has the advantage of avoiding confounding due to unwanted variations in th ocular system that may be associated with vari ations of the degree of optical transformation

The model of adaptation to uniocular imag magnification which we have advanced in our earlier reports (Epstein, 1971, 1972a; Epstein & Morgan, 1970), and which has been ind pendently formulated by Wallach (Wallach & Frey, 1972a) in accounting for adaptation in volving the oculomotor cues, has a ready explanation for the dependence of magnitude of adaptation on degree of informational discrepancy. In the present instance, the claim is that binocular disparity has been recalibrated or reevaluated as a consequence of being continuously paired with a dominant, discrepant input. The degree of disparity imposed by the MSL acquires a new significance for perceived depth, one which is consistent with the dominant perspectival slant. The farther apart are the two slants, the greater must be the shift in the significance of disparity if the prevailing degree of disparity is to become equivalent in its effect on perceived depth to the effect of perspective.

It is important to note that in our formulation, the directing factor in the relationship between the discrepant inputs is not that one elicits veridical percepts and the other, being transformed, elicits nonveridical percepts. In the present experiment, since the target was in fact frontoparallel, both inputs induced non-

veridical percepts. The directing factor in our experiment was cue dominance. As Experi[illegible] showed, under our conditions per[illegible]al slant was dominant. Accordingly, [illegible]ity was recalibrated. The direction of [illegible]ift, therefore, is determined by dominance, not veridicality. Of course, if the dominant input is untransformed and yields veridical [illegible]on the adaptive shift in the subordinate [illegible]ll be toward veridicality. That shifts [illegible]eridicality have been the rule in the [illegible]f adaptation is a coincidence of the [illegible]ental designs that have been favored. [illegible]imple matter to demonstrate "counter[illegible]ptation" (Wallach & Frey, 1972b) or "maladaptive adaptation" (Canon, 1971), in which the shift is from veridical to nonveridical perception. All that is required is to pair a subordinate nontransformed input with a dominant transformed input.

The vital role of cue dominance as the directive factor in adaptation may also explain the way in which selective attention effects adaptation to informational discrepancy. Canon (1970) exposed S to an intermodal discrepancy between visual and auditory localization. During the adaptation period, S tracked a moving target presented in both modalities, under instructions to attend to the visual input ignor[illegible] the auditory input, or vice versa. Follow[illegible]he adaptation period, measures of locali[illegible]tion were secured with auditory input and visual inputs separately. Canon found that adaptation was restricted almost entirely to the unattended modality. If we may assume that selectively attending to a cue or to the input from a specific modality confers dominance to the attended cue or modality, Canon's finding can be understood. Dominance relationships among cues or modalities are no doubt well-established by preexperimental perceptual experience, perhaps on the basis of probabilistic considerations similar to those proposed by Brunswik (1956). But dominance in a given situation may be affected by contemporary or situation-specific factors, in Canon's (1970) case, distribution of attention. This result is all the more significant because it seems likely that without selective attention visual localization dominates auditory localization (Weerts & Thurlow, 1971), so that if our interpretation is correct, selective attention is able to reverse preexperimental dominance relationships.

The 60° inspection figure in Experiment II was selected because in this case perspectival slant and stereoscopic slant matched. Accordingly, informational discrepancy was assumed to be absent and significant adaptation was not expected. However, the experimental results showed that although the 60° inspection figure yielded the smallest shift the adaptation effect was significant. This result may suggest that a portion of the adaptive shift is not a response to informational discrepancy, just as in the case of tilt adaptation a part of the total adaptive shift may be attributed to Gibson-type normalization (Morant & Beller, 1965). A more likely explanation of the occurrence of significant adaptation to the 60° figure is that motion parallax, arising from unavoidable small head movements during the 10-min. adaptation period, offset perspectival slant and introduced a discrepancy between motion parallax slant and stereoscopic slant. This difficulty would not be present in Experiment I because S was not required to maintain the head in a stationary position for any extended time. When we retested eight high adapters selected from the Ss of Experiment II some time after the conclusion of Experiment II, using a modified procedure which reduced head movements to the practical minimum, the adaptive shift in response to the 60° figure was reduced. Nevertheless, since adaptive shifts were not totally eliminated the question remains open.

The other principal finding of Experiment II is that an adaptation activity that requires depth processing yields greater adaptation than passive registration of depth. The effect of processing level may be attributable to two related factors: (*a*) depth processing may supply an added motive for resolution of the informational discrepancy, inasmuch as the requirement to localize the target may compel the visual system to opt for one outcome, and (*b*) depth processing may enhance the potency of the dominant cue by encouraging S to utilize the cue. Although the experimental results clearly show an enhancement attributable to depth processing it is also clear that adaptation occurs even when S is directed away from the relevant discriminations. Of course, the procedure did not purport to control S's processing completely, only to introduce a difference in level of processing. It seems likely that in the depth-registration condition Ss did engage in a degree of depth processing.

It should be understood that we are not claiming that depth processing is a necessary condition for the occurrence of adaptation to transformed disparity. Even were depth processing absent completely, the fact that dis-

crepant inputs are present is likely to be sufficient to produce some adaptation. The situation is analogous to that which prevails in the case of the difference between passive and active movement. While adaptation usually is greater following active movement, significant adaptive shifts do occur following passive movement (e.g., Mack, 1967; Singer & Day, 1966). According to our argument, adaptation is greater following active self-generated movement for reasons similar to those we offered in explaining the effect of level of processing in the present experiment. Nevertheless, adaptation following passive movement will occur as long as informational discrepancy is present.

REFERENCES

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: University of California Press, 1956.

Canon, L. K. Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, 1970, **84**, 141–147.

Canon, L. K. Directed attention and maladaptive "adaptation" to displacement of the visual field. *Journal of Experimental Psychology*, 1971, **88**, 403–408.

Clark, W. C., Smith, A. H., & Rabe, A. Retinal gradient of outline as a stimulus for slant. *Canadian Journal of Psychology*, 1955, **9**, 247–253.

Ebenholtz, S. M. Adaptation to a rotated visual field as a function of degree of optical tilt and exposure time. *Journal of Experimental Psychology*, 1966, **72**, 629–634.

Efstathiou, E. Effects of exposure time and magnitude of prism transform on eye-hand coordination. *Journal of Experimental Psychology*, 1969, **81**, 235–240.

Epstein, W. *Varieties of perceptual learning.* New York: McGraw-Hill, 1967.

Epstein, W. Adaptation to uniocular image magnification after varying preadaptation activities. *American Journal of Psychology*, 1971, **84**, 66–74.

Epstein, W. Adaptation of uniocular image magnification: Is the underlying shi[illegible] rioceptive? *Perception & Psychophysics*, 1972, **11** 89–91. (a)

Epstein, W. Retention of adaptat[illegible] to uniocular image magnification: Effect [illegible] interpolated activity. *Journal of Experime[illegible] Psychology*, 1972, **92**, 319–324. (b)

Epstein, W., & Daviess, N. Modification of depth judgment following exposure to magnification of uniocular image: Are changes in perceived absolute distance and registered direction of gaze involved? *Perception & Psychophysics*, 1972, **12**, 315–317.

Epstein, W., & Morgan, C. L. [illegible]tation to uniocular image magnification: [illegible]tion of the disparity-depth relationship. [illegible]can Journal of Psychology*, 1970, **83**, 322–329.

Freeman, R. B., Jr. Absolute th[illegible]old for visual slant: The effect of stimulus siz[illegible] retinal perspective. *Journal of Experim[illegible]l Psychology*, 1966, **71**, 170–176.

Gillam, B. Perception of slant when perspective and stereopsis conflict. *Journal of Experimental Psychology*, 1968, **78**, 299–305.

Mack, A. The role of movement [illegible]erceptual adaptation to a tilted retinal ima[illegible] *Perception & Psychophysics*, 1967, **2**, 65–68.

Morant, R. B., & Beller, H. K [illegible]daptation to prismatically rotated visual fields. *Science*, 1965, **148**, 530–531.

Ogle, K. N. *Researches in binocular vision.* New York: Hafner, 1964.

Seagrim, G. M. Stereoscopic vision and aniseikonic lenses: Further studies in latency. *British Journal of Psychology*, 1969, **60**, 307–313.

Singer, G., & Day, R. H. Spatial adaptation and aftereffect with optically transformed vision: Effects of active and passive responding and the relationship between test and exposure responses. *Journal of Experimental Psychology*, 1966, **71**, 725–731.

Wallach, H., & Frey, K. J. Adaptation in distance perception based on oculomotor cues. *Perception & Psychophysics*, 1972, **11**, 77–83. (a)

Wallach, H., & Frey, K. J. On counteradaptation. *Perception & Psychophysics*, 1972, **11**, 161–165. (b)

Weerts, T. C., & Thurlow, W. R. The effects of eye position and expectation on sound localization. *Perception & Psychophysics*, 1971, **9**, 35–39.

(Received May 14, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 595–603

EVIDENCE FOR SEMANTIC ANALYSIS OF UNATTENDED VERBAL ITEMS[1]

MARILYN C. SMITH[2] AND MARY GROEN

Scarborough College, University of Toronto, Toronto, Ontario, Canada

Are unattended verbal materials stored in short-term memory solely in terms of their sensory features, or do they receive some semantic analysis? To answer this question, *S*s received dichotic presentation of lists of words, with the 2 lists being selected either from the same or from different categories. The *S*s were instructed to attend to only one ear, and ignore the other. Following list presentation, a probe was presented, with *S*s indicating if that probe had been presented in the attended list. Probe type was varied as to category similarity with the attended or unattended lists. It was found that when a probe item was presented that had been an actual item in the unattended channel, long latency and high error rates were found if the category of the unattended item was the same as the attended list, but not if they differed. This was interpreted as indicating that unattended materials receive some semantic analysis.

Several recent studies have pointed to the existence of short-term memory (STM) storage for unattended verbal items (Davis & Smith, 1972; Norman, 1969; Peterson & Kroener, 1964). However, there is some ambiguity concerning the extent to which this material is processed prior to storage. Glucksberg and Cowan (1970) argued that it is only the sensory features of the stimulus, rather than semantic or categorical information about it which is held in STM. This conclusion was based upon 2 findings. In a shadowing task, where narrative passages were dichotically presented, digits were occasionally presented to the unattended ear. When *S*s were later asked if a digit had been presented, homophones of digits that occurred in the narrative such as "too" and "for" were incorrectly identified as digits. This indicates that context had no effect, and that information about meaning was therefore not being extracted from the unattended channel. Secondly, *S*s were instructed to report if there were any instances where they believed a digit had been presented, but were unable to recall exactly which digit it was. However, it was found that unless a specific digit was reported, *S*s reported that no digit had occurred. There was thus no apparent memory for the category *digit* that was separate from specific item occurrence. This is in contrast to memory-attended events, where *S*s can report, correctly, that a category member (e.g., a digit or a word) has occurred without being able to report the identity of the specific item.

Some data of Treisman (1964) support this contention. When an unattended message was the same as the one being shadowed, all *S*s noticed this identity when the time separation was sufficiently short. The critical time interval depended upon whether the shadowed or rejected message was leading. When the relevant message was leading, identity was noticed with an interval up to 4.5 sec., whereas when the irrelevant message was leading, it was only 1.4 sec. Treisman considered these durations to be a measure of STM for the 2 kinds of materials. The important finding with regard to the question of semantic analysis of unattended input relates to the differences she found in the memory duration of attended and unattended messages as a function of the nature of the material employed. When the relevant message was leading, the interval for noticing

[1] This research was supported by National Research Council of Canada Grant APA-191 to Marilyn C. Smith.

[2] Requests for reprints should be sent to Marilyn C. Smith, Department of Psychology, Scarborough College, University of Toronto, Toronto M5S 1A1, Canada.

identity was 4.5 sec. for prose. This corresponds to about 12 words, the usual memory span for this type of material. However, when *S* was shadowing isolated words, the critical lag was much smaller, about 6 words. Thus, memory in the attended channel is a function of meaning. On the other hand, the number of words by which the irrelevant message led when identity was noticed did *not* depend upon the kind of material used at all, suggesting that semantic analysis was not occurring.

Other studies, however, provide evidence that some semantic analysis of unattended events is done. Lewis (1970) found that unattended material that was semantically similar to the attended material (synonyms) caused *S*s more difficulty in a shadowing task than did semantically dissimilar items. Similarly, Davis and Smith (1972) found that the dichotic presentation of a list of nonsense words which *S*s were instructed to ignore did not cause as much interference in the recall of a list of attended words as did the presentation of a to-be-ignored list of real words.

In the study reported here, a reaction time (RT) paradigm introduced by Sternberg (1966) was employed to study the question of semantic analysis of unattended material. In this procedure, *S*s are presented with a list of words to remember (memory set) followed by a probe item. The task is to determine whether the probe item had been presented as a member of the memory set. It generally has been found that RT increases linearly with the size of the memory set, and that the slope of the function is the same whether or not the probe was one of the presented items. Sternberg interpreted these results as indicating that *S*s perform a serial exhaustive scan of the items in memory.

Recent elaboration of this procedure has indicated that in some situations *S*s may be capable of categorizing the items in the memory set so as to reduce the required search time. De Rosa and Morin (1970) found that when the memory set consisted of a set of consecutive digits, the time required to determine that the probe digit was not a member of this set increased as the difference between the arithmetic value of the presented stimulus and the highest (or lowest) digit in the memory set increased. Lively and Sanford (1972), presenting lists consisting either of digits or of letters, found that the time to decide that a probe item was not a member of the memory set was faster if the probe was a member of the other category to that employed in the memory set. Further evidence for such categorization of items is presented by Smith and Abel (1973). They presented a list of items drawn from several categories followed by a probe consisting of a category name. It was found that the time required to decide whether the given category had been represented in the memory set varied directly with the number of categories, rather than with the actual number of items. All of these studies suggest that upon presentation of the memory set, some categorization of the list items occurs.

The above experiments employed only memory sets to which *S*s were attending. Does the same kind of category analysis occur for unattended input? To study this, *S*s were dichotically presented with 2 lists of words, one of which they were instructed to ignore. All items within a list were drawn from the same category. By varying whether the unattended list on a given trial contained items belonging to the same category as the attended list, as well as the category from which the probe item was selected, it was hoped that something could be learned about the nature of the semantic analysis of unattended input.

If unattended materials are not semantically analyzed, but have only acoustic representation in STM, as suggested by Glucksberg and Cowan (1970), then category similarity between attended and unattended lists should have no influence on RT to the probe. Any effect of category similarity, on the other hand, would be taken as evidence that some semantic analysis is done of the ignored channel. To get more information about the memory representation for unattended input, several different probe words were employed.

Positive probes (i.e., those requiring a *yes* response) were items that had actually been presented in the attended channel. Negative probes were selected from (*a*) the category employed in the attended list, (*b*) the category employed in the unattended list, (*c*) a different category entirely, or (*d*) an actual item in the unattended list. A monaural condition in which the category of the negative probe could be the same or different from the memory set was included for comparison.

Method

Subjects. Five male university students with normal hearing volunteered to participate in this experiment, and were paid for their services.

Apparatus. All lists were recorded using a magnetic Sony (TC-355) 4-track tape recorder, and were presented individually to *S*s through headphones. In recording the lists for dichotic presentation, the 2 channels were matched approximately for loudness. By pressing a voice-operated relay key (Uhr Model F422) before recording each probe word, an inaudible mark was made on the second track of the tape at the onset of the probe. There was a variable delay of 20–40 msec. between actual probe onset and the mark on the tape. During playback, a sensing device, on detecting this mark, triggered a Hunter Klockounter (Model 120A, Series D) which measured RTs to the nearest millisecond. The timer was stopped when *S* manually depressed 1 of 2 microswitch response buttons to indicate his response choice. Because of the distance on the tape recorder between the recording and playback heads, the probe word was not actually heard to begin until 256 msec. after the tape mark had triggered the timer. Thus, this time was subtracted from all recorded RT measures.

Procedure. Six hundred 2-syllable words from 45 different categories were selected from Battig and Montague's (1969) category norms. Only those list words that had been included as category members by at least 5 of the students who established the norms were included. No word was included in more than one category. Prior to testing, *E* read *S* a list of all the words to be used to ensure familiarity with all words and pronunciations.

Each *S* attended 7 sessions over a 4-day period. Two sessions were presented on each of Days 1–3, and the last session was presented on Day 4. On Day 1, *S* received only monaural presentation of the lists, the first session representing a practice session and the second the test. Dichotic lists were presented on Days 2–4, and consisted of 1 practice session followed by 4 test sessions.

In each session, 144 trials were presented. A single trial consisted of the following sequence: a warning signal (metallic tap), a list of 2, 4, or 6 words, another warning signal, and finally, the probe word. A 6-sec. interval elapsed before the beginning of the next trial. In the monaural trials, words and signals occurred at the steady rate of 1/sec. For dichotic trials, one pair of words per second was heard, one word from each tape channel per second. Words in a pair were approximately simultaneous, with the onset of the attended word always preceding the onset of the ignored word by 100–150 msec. Half of the probes were positive; i.e., they were items that had occurred in the preceding list and required a *yes* response. The other half of the probes were negative, requiring *S* to respond *no*. Positive probes were selected from each serial position with equal frequency.

For each *S*, one ear was chosen as the "relevant" ear. Instructions emphasized that he was to attend *only* to that ear, and to ignore materials presented to the irrelevant ear. The *S*s were warned that some probes would occasionally be drawn from the list presented to the irrelevant ear, but that these required a *no* response, and that they would do better in the task if they totally ignored the irrelevant materials. In an attempt to aid *S*s in blocking out the irrelevant ear, the monaural and dichotic attended lists were always presented to the same ear. Further, in the dichotic condition the warning signals and probe words were presented only in the relevant ear. The *S* always made his positive response with the same hand. Each of the 4 possible ear–hand combinations was assigned to each of 4 *S*s. One of the combinations was then randomly assigned to the remaining *S*.

Words within a single list represented only one category. For each trial, a category was selected at random, and then the appropriate number of words was randomly chosen from that category without replacement in a single session. A given word was used only once in the monaural test session, but twice in each dichotic session, once in an attended list and once in an unattended list.

The number of trials in each condition differed for monaural and dichotic sessions. In monaural sessions, half of the negative probes were drawn from the same category as the list items, and half from a different category. Dichotic trials were first subdivided according to the categories of the 2 lists. On half of the trials the lists presented to each ear were from the same (S) category; on the other half, they were from different (D) categories. Fifty percent of the probes were positive. The remaining 50% of the negative probes were subdivided into the following types: (*a*) a member of the same category as the attended list (Types S_1 and D_1); (*b*) an actual item from the ignored list (Types S_2 and D_2); (*c*) an item from a category different from both lists (Types S_3 and D_3); or (*d*) if the 2 lists were from different categories, an item from the same category as the unattended list (Type D_4).

Under Condition S, the first 3 negative probe types were tested with equal frequency. For Condition D, all 4 types were used, again with equal frequency. Samples of each type of probe appear in Tables 1 and 2, together with their frequency of presentation in each session. If the

TABLE 1

Examples of Probe Types for Monaural Condition (List Length

List item	Probe	Type	Description	Frequency
GIRAFFE	GIRAFFE	Positive	Serial Position 1	24[a]
LION	TIGER	Negative	Same category as list	12
ZEBRA	LEMON	Negative	Different category	12
MONKEY				
Total				48[b]

[a] This figure derived from 4 Serial Position × 6 Replications.
[b] The total of 144 was derived from 48 Total Frequency × 3 List Lengths.

negative probe was an actual item from the ignored list, each serial position was tested equally often.

The specific order of conditions within a session was randomized separately for each set of 144 trials. Each session was recorded on a separate tape. Four dichotic test tapes were required in order to test each serial position for S_2 and D_2 probes at least twice for every list length. They were presented in a different random order to each S.

Results

Monaural presentation. A 3 × 3 × 5 analysis of variance was performed on the mean RT data, the factors being list length (2, 4, or 6 words), probe type (positive, negative from the same category as list, and negative from a different category), and *S*s. Incorrect responses were not included in the analysis. The only significant factor found was list length, $F\ (2, 8) = 26.7$, $p < .01$. Probe types, the variable of interest with respect to classification, did not result in significantly different RTs, suggesting that classification was not used as the decision strategy in responding to negative probes. Despite this finding, all *S*s spontaneously reported that probes from different categories were easier to reject.

Dichotic presentation. Mean RTs at each list length as a function of whether the category of the attended and unattended lists are the same or different are shown

TABLE 2

Examples of Probe Types for Dichotic Condition (List Length 4)

List: Attended	List: Ignored	Probe	Type	Description	Frequency
				Two lists from same (S) category	
JOANNE	EILEEN	BETTY	Positive	Serial Position 2	12[a]
BETTY	SUSAN	MARY	Negative	Same category as attended (S_1)	4
SHARON	LOIS	CAROL	Negative	Item from ignored list, Serial Position 4 (S_2)	4
NANCY	CAROL	SQUIRREL	Negative	Different category (S_3)	4
Total					24[b]
				Two lists from different (D) category	
APPLE	BANJO	LEMON	Positive	Serial Position 3	12
ORANGE	CELLO	MANGO	Negative	Same category as attended (D_1)	3
LEMON	TUBA	CELLO	Negative	Item from ignored list, Serial Position 2 (D_2)	3
CHERRY	FIDDLE	TIGER	Negative	Different category (D_3)	3
		GUITAR	Negative	Same category as ignored list (D_4)	3
Total					24[b]
Grand total					144[c]

[a] This figure was derived from 4 Serial Position × 3 Replications.
[b] A total of 72 was derived from 24 Total Frequency × 3 List Lengths.
[c] The total of 144 was derived from 72 Total from Same Category + 72 Total from Different Category.

in Figure 1. Monaural RT data, in which positive and negative probes have been combined, are also shown for comparison. A 4-way analysis of variance was performed on correct mean dichotic RT measures, the factors being list length (3); probe type, i.e., positive or negative (2); category classification, i.e., attended and unattended lists from the same or different categories (2); and *Ss* (5). Different types of negative probes were combined for this preliminary analysis. Only the effects of list length and category type were significant ($p < .01$). As lists became longer, RT increased, F (2, 8) = 70.6, and responses to probes were made more slowly when both lists were from the same category than when the categories were different, F (1, 4) = 26.7. None of the interactions reached the .05 level. A replication of the analysis of variance using median scores yielded the same results.

It can also be seen in Figure 1 that monaural RT was faster than dichotic RT at all list lengths. This effect did not reach significance at the .05 level, perhaps because monaural trials always preceded dichotic trials, and *Ss* were thus not as well practiced. Least squares linear regression equations were calculated for the 3 functions. They are shown in Figure 1. The slope for the 2 dichotic conditions, 44.7 msec. per list item for same category lists and 41.4 msec. for different category lists, indicate that the 2 functions are essentially parallel. Both are slightly steeper than the monaural condition (33.7 msec. per item).

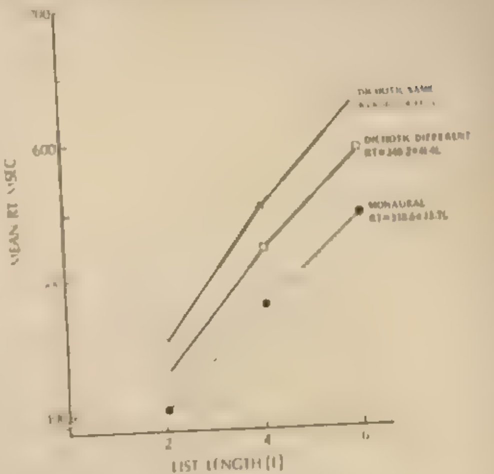


FIGURE 1. Reaction time (RT) for monaural and dichotic conditions.

Mean correct RT for each of the 7 negative probe types as a function of list length is shown in Figure 2. An independent analysis of variance of the various types of negative probes was carried out to clarify the interpretation of the observed differences. The S_2 probes, i.e., those in which an actual item from the unattended list was presented under the condition where both lists were from the same category, were not included in the analysis because of the large number of errors that occurred in this condition. The remaining 6 probe types were examined in a Probe Type × List Length × *Ss* factorial analysis of variance design. Both probe type, F (5, 20) = 5.5, $p < .01$, and list length, F (2, 8) = 21.5, $p < .01$, were found to be significant variables. For this analysis, only correct responses of less than 2 sec. latency were included in determining means.

All sets of paired comparisons of probe types were made by the least significant difference test. When a .05 level was assumed, the 6 probe types fell into 2 different groups. No differences were found for S_1, D_1, and D_2 items. These 3 types were all responded to more slowly than S_3, D_3, and D_4 probes, with the latter 3 not being differently responded to.

To summarize the data in Figure 2, the slowest negative RTs were made to actual items from the unattended list in the condition where both lists were drawn from the same category. Collapsing across list length, mean RT for the S_2 condition was 713 msec. The second longest RTs were found when a probe was presented from the unattended list, the 2 lists being drawn from different categories (D_2), and in those conditions where the negative probe was from the same category as the attended list, regardless of category similarity of the 2 lists (S_1 and D_1). The mean RTs for these 3 conditions were 544, 534, and 547

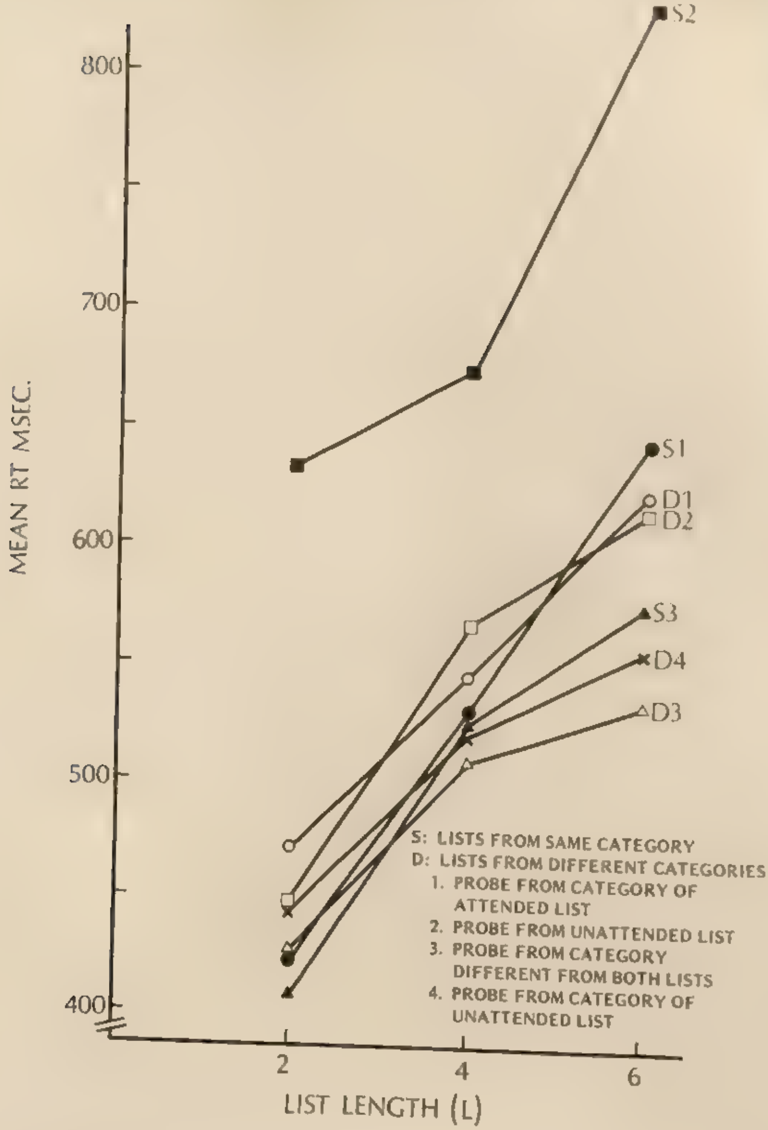


FIGURE 2. Negative reaction times in dichotic condition for each probe type.

msec., respectively. Finally, the fastest negative responses occurred in the 3 conditions in which the probe was drawn from a different category from the attended list, even if that item represented the category of the unattended list, providing it was not an actual item from the unattended list. The mean RTs for these 3 conditions were 505 (S_3), 491 (D_3), and 508 (D_4) msec.

Negative probes that represented actual items from the unattended list (S_2 and D_2) were examined further for serial position effects. Plotting the mean RT as a function of serial position for each list length did not reveal any consistent trends.

Errors. The mean percent errors at each list length are shown in Figure 3. Positive and negative responses have been combined in the monaural condition. Error rate in the monaural condition did not vary as a function of list length, and averaged 3.1%. With dichotic presentation, the errors for negative probes with the exception of the

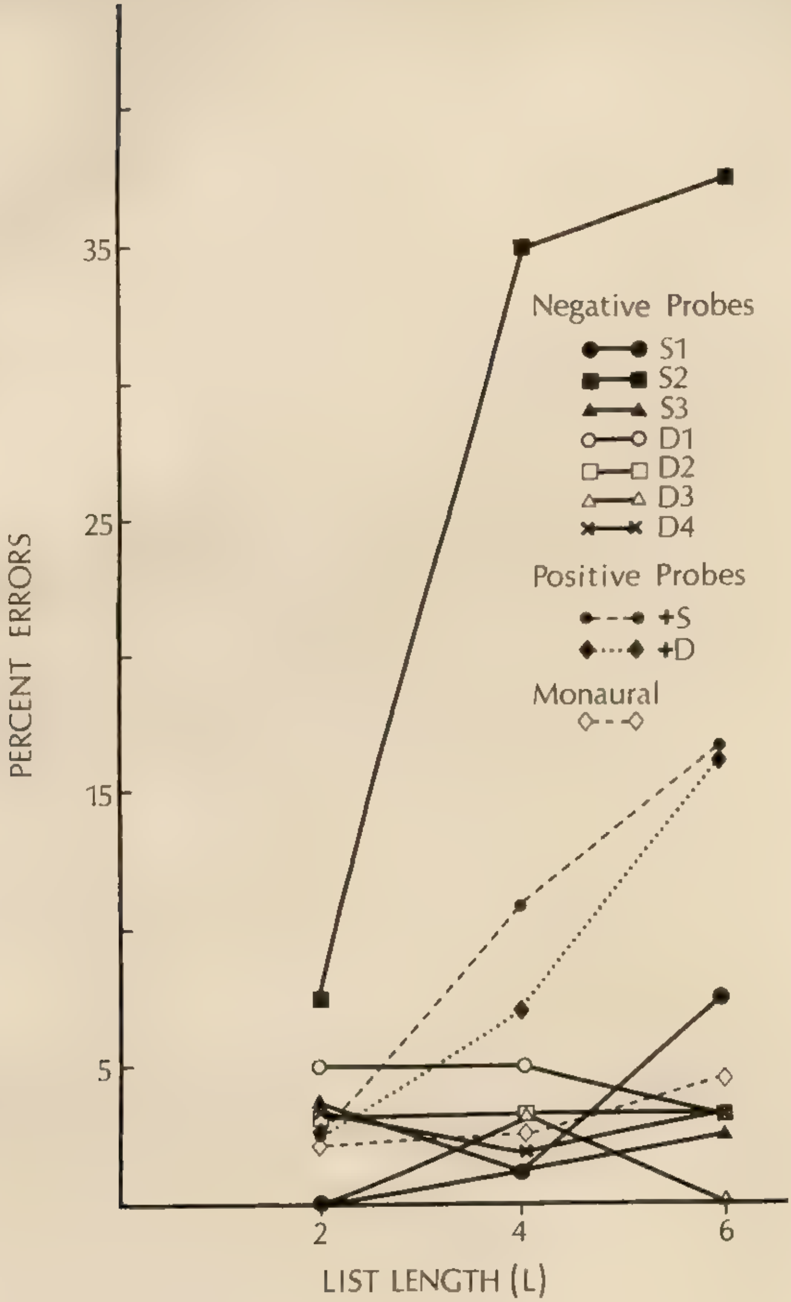


FIGURE 3. Mean percent errors for monaural and dichotic conditions. (Abbreviations: S = same, D = different.)

S_2 condition, were the same as for monaural presentation, in that they did not vary across list length, and averaged 2.9%. For positive responses in the dichotic condition, errors were found to be greater than for negative probes, and to increase with increases in list length.

By far the greatest number of errors occurred in the S_2 condition, where the probe is an actual item from the unattended

list, both lists being drawn from the same category. For List Length 2, error rate was a relatively small 7.5%. However, for the longer list lengths it was approximately 35%, indicating that *S*s have great difficulty distinguishing these items from items in the attended list. In contrast, only 2.8% of the responses to D_2 probes were errors. To determine whether the error rate for the S_2 probe was a function of serial position in the unattended list, error rate was plotted as a function of serial position for each list length. However, no consistent pattern was found.

Serial position effects for positive probes. Mean RT and percent errors for positive probes as a function of serial position were examined for both the monaural and various dichotic conditions. For List Length 2, RT and errors were greatest in Serial Position 2. For List Length 4, a bow-shaped function was found for both RT and errors, with slowest responses and greatest number of errors occurring toward the middle of the list. Finally, at List Length 6, while the RT curve was again generally bowed for the monaural condition, a marked recency effect was found in the dichotic conditions. Similarly , errors were greatest for the items at the beginning of the list, with fewest errors occurring at the later serial positions.

Discussion

The results of this experiment provide evidence that the unattended input is not stored solely in terms of sensory features. Rather, some category analysis appears to take place. The strongest evidence for this conclusion comes from the 2 conditions in which an actual item is drawn from the irrelevant channel and presented as a probe (requiring a negative response). If all unattended items are stored as acoustic representations, as suggested by Glucksberg and Cowan (1970), the category membership of the unattended message should be unimportant. However, it was found that if the items in the unattended channel belonged to the same category as those in the attended list (S_2), RT was far greater than in the condition where the unattended message belonged to a different category (D_2).

The 2 conditions also differed greatly with respect to errors. If a probe, selected from the unattended channel, belonged to a different category from the attended list (D_2), error rate was low, resembling that found for the monaural condition, and did not vary with list length. On the other hand, if the probe was an item that had been presented in the ignored ear and did belong to the same category (S_2), error rate was very high, increasing sharply with list length.

It is of interest to note that the effect of irrelevant list category on either RT or errors is found only when the probe is an actual item from the unattended list. A negative probe from the same category as the unattended list (D_4) is not responded to differently than a probe from a category that was not presented at all (D_3). Hence, category membership of the unattended items is only an important variable when the sensory or acoustic representation of the items are actually present. This would lend support to the conclusion of Glucksberg and Cowan (1970) that there does not appear to be any abstract category representation for unattended items.

The finding that negative probes from the same category as the attended message (S_1 and D_1) are rejected more slowly than words from a different category (S_3, D_3, and D_4), is in agreement with the results of Lively and Sanford (1972). It indicates that categorization occurs for the attended message, with this information being used as a basis for rejection. The failure to find this effect in the monaural condition (i.e., shorter latency for noncategory negative probes) may be due to the fact that unlike the Lively and Sanford study, where only 2 highly learned categories were used (digits and letters), many different categories were employed in this experiment. Since the monaural condition was always presented first, *S*s were less practiced than in the dichotic condition, and perhaps registration of category information was sufficiently variable to eliminate any differences.

References

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt. 2).

Davis, J., & Smith, M. C. Memory for unattended input. *Journal of Experimental Psychology*, 1972, **96**, 380-388.

De Rosa, D. V., & Morin, R. E. Recognition reaction time for digits in consecutive and nonconsecutive memorized sets. *Journal of Experimental Psychology*, 1970, **83**, 472-479.

Glucksberg, S., & Cowan, G. N. Memory for nonattended auditory material. *Cognitive Psychology*, 1970, **1**, 149–156.

Lewis, J. L. Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 1970, **85**, 225–228.

Lively, B. L., & Sanford, B. J. The use of category information in a memory-search task. *Journal of Experimental Psychology*, 1972, **93**, 370–385.

Norman, D. A. Memory while shadowing. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 85–93.

Petersen, L. R., & Kroener, S. Dichotic stimulation and retention. *Journal of Experimental Psychology*, 1964, **68**, 125–130.

Smith, M. C., & Abel, S. Memory search: When does semantic analysis occur? *Perception & Psychophysics*, 1973, **13**, 233–237.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

Treisman, A. M. Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 449–459.

(Received May 25, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 604–608

FEEDBACK PRECISION AND POSTFEEDBACK INTERVAL DURATION[1]

CECIL A. ROGERS, JR.[2]

University of Arizona

Precision of feedback grain was manipulated in a simple positioning t An optimum was found; an increase in precision past that optimum proc deleterious effects upon rate of acquisition. In a second study, increasing postfeedback interval removed that optimum. The feedback precision eff were then replicated in a timing task. The combined results of the 3 stud were interpreted as supportive of an information-processing approach to t study of postfeedback interval events for simple motor skills. The findi additionally supported specific predictions by Bilodeau and deductions Adams' 1971 theory of motor learning.

Trowbridge and Cason (1932) provided early evidence that increases in feedback precision normally promote rate of acquisition in perceptual–motor tasks. Reviews of the literature by Adams (1971) and Bilodeau (1966, 1969) confirm this relationship for a variety of tasks. Bilodeau did suggest, however, that for any specific task, some optimal level of feedback grain might exist. Increases in precision beyond that level might fail to enhance and could be detrimental to the learning process. Extrinsic feedback might be refined past the point of human response capability or could exceed *S*'s information-processing capacity.

One previous study by Bilodeau and Rosenbach (1953) found few effects from varying feedback grain in a knob positioning task. Since their range of manipulation of feedback precision was somewhat narrow, Experiment I was designed to investigate broader variations of feedback precision in the same task.

The first study was, in essence, a search for Bilodeau's (1966, 1969) predicted optimum. The task and procedures were adapted from Bilodeau and Rosenbach (1953). Their micrometer positioning task met criteria that the amount of intrinsic feedback should be relatively small and that response error measurement should be finely grained.

Experiment I

Method

Subjects. The *S*s were 80 male and female volunteers from introductory psychology classes at the University of Arizona. They were randomly assigned to 1 of 4 groups with the restriction that each group contain equal numbers of each sex.

Apparatus and procedure. A micrometer (Starrett Model 263RL) was mounted behind a black wooden screen so that its turning knob projected through a hole in the center of the screen (Bilodeau & Rosenbach, 1953). The *S*s were seated and their arms were positioned perpendicular to the apparatus, which rested upon a table. The view presented to *S* was that of a black knob on a black background. The body of the micrometer rested behind the screen and was visible only to *E*. Verbal instructions to *S* indicated that he must rotate the knob and learn to stop it at the correct position. One complete rotation of the knob produced a micrometer shaft displacement (not visible to *S*) of .250 unit; the goal response was set at 4 units ± .5 unit. Response error was read directly from the micrometer by *E* before rounding and reporting it orally to *S*. Error was recorded to the nearest .001 unit for all groups. The micrometer was calibrated such that 1 unit equaled 1 in.

Ten acquisition trials were given to each *S*. A single trial consisted of a verbal ready signal followed in 2 sec. by a go signal which was followed in turn by, *S*'s response, oral feedback 43 sec. after the go signal, and the next ready signal 5 sec. after termination of feedback. Duration of feedback was approximately 4 sec. and the total trial cycle was about 55 sec.

[1] This research was supported in part by National Aeronautics and Space Administration Grant NGR-03-002-091 and in part by Tulane University while the author was on sabbatical leave there, 1971–72. The assistance of I. McD. Bilodeau and the data collection aid of John Ireland are appreciatively acknowledged.

[2] Requests for reprints should be sent to Cecil A. Rogers, Jr., Department of Psychology, University of Arizona, Tucson, Arizona 85721.

On the first trial, all *Ss* were told to turn the knob as far as they cared to. If *S* hit the target on a subsequent trial, he was told that his response was correct and that he should try to repeat it on the next trial.

Design. The *Ss* in the 0-order treatment were informed of direction of response error relative to the goal (too short or too far). The 1-digit group also received magnitude information rounded to the nearest whole unit. Response error for 2- and 4-digit groups was reported to the nearest tenth and thousandth of a unit, respectively. For example, an actual response error of +3.214 in. would be reported as "too far," +3, +3.2, or +3.214 units, depending upon feedback condition.

Results and Discussion

Absolute error data were consolidated into 5 2-trial blocks as displayed in Table 1. Variance analyses of Block 1 revealed no significant differences among groups at the onset of practice, F (3, 76) = .40. Subsequent analyses revealed a significant practice effect, F (4, 304) = 72.38, $p < .01$, and a significant feedback grain effect, F (3, 76) = 3.41, $p < .025$. No Feedback × Practice interaction was indicated, F (12, 304) = 1.62. As shown in Table 1, mean response error showed progressive diminution with practice and the increase of feedback precision from 0 to 1 to 2 digits of information. The 4-digit group showed a reversal of this trend by performing more nearly like the 0-digit condition. Examination of Block 5 data revealed a feedback precision effect; F (3, 76) = 3.83, $p < .025$. An unequal interval trend analysis of precision levels at Block 5 indicated a curvilinear component, F (1, 76) = 11.0, $p < .01$. Bilodeau's (1966, 1969) hypothesis was then confirmed in that the most precise level of feedback administered in Experiment I not only failed to produce optimal acquisition rates, but slowed acquisition relative to the 2 intermediate feedback levels.

An analysis of within-trial events suggested a possible mechanism for operation of the feedback grain effects found in Experiment I. Postfeedback interval, from termination of feedback to the next response, was 7 sec. Adams (1971) reviewed the effects of interpolating motor responses in the postfeedback interval and

TABLE 1

Mean Response Error (in in.) for Feedback Conditions over Trial Blocks: Experiment I

Digits of feedback	Trial block				
	1	2	3	4	5
0	5.92	4.91	4.47	3.18	3.14
1	5.86	4.05	2.90	2.33	1.92
2	5.45	4.05	2.63	1.26	.87
4	6.03	4.81	4.17	3.85	2.81

found few effects on motor-skill acquisition. He hypothesized that the interval is primarily one of verbal–cognitive activity by *S*, wherein feedback is processed and a response modification is computed. Adams further suggested that acquisition might be slowed either by interpolating verbal activity in the postfeedback interval or by excessively foreshortening it. If some such minimal information-processing time is required between feedback and the next response, it seems likely that feedback precision would also in part determine optimal interval duration. As precision increases, more information is imparted to *S*. As Adams notes, such precision normally provides *S* with information to make more finely graded responses. However, as *S*'s informational load is increased, more time is required for his selection of a response modification. For a given task, acquisition might then be hindered by very precise feedback combined with a short postfeedback interval. The 7-sec. postfeedback interval in the present study may have been inadequate for the 4-digit condition. The present results considered in the context of Adams' theory suggest a possible interaction between feedback precision and postfeedback interval duration.

Weinburg, Guy, and Tupper (1964) previously demonstrated the deleterious effects of shortening the postfeedback interval while holding grain constant. For one positioning task, they found that 5 sec. was the minimal period required between feedback and the next response. While increases beyond 5 sec. did not aid acquisition, reduction below 5 sec. was harmful.

Experiment II was then designed as a replication of Experiment I, but with the postfeedback interval lengthened. An improvement in the ranking of 4-digit *Ss* might then be expected in the context of Adams' (1971) theory. Bilodeau's (1969, p. 277) position, which emphasizes the information-processing aspects of the postfeedback interval, would also be suggestive of such a finding. Such an outcome would further be consistent with data obtained on postfeedback intervals by Bourne and Bunderson (1963) for concept formation tasks.

Experiment II

Method

Subjects. Sixty undergraduate volunteers from Tulane University were divided into 3 groups of 20 *Ss*. Each group was comprised of an equal number of males and females.

Apparatus and procedure. No change was made in apparatus. Procedure differed from Experiment I only in lengthening of postfeedback interval from 7 to 14 sec., with a consequent foreshortening of the prefeedback period from 43 sec. to 36 sec. The total trial cycle remained 55 sec.

Design. The 0-digit feedback condition present in the first study was omitted from Experiment II. Feedback grains of 1, 2, and 4 digits were administered as in Experiment I to the 3 remaining groups over 10 trials.

Results and Discussion

Mean absolute response error per 2-trial block is shown in Table 2 for the 3 feedback grain conditions. A reordering of rank-order positions obtained in comparison to those found in Experiment I. The relative performance levels of the 1- and 2-digit conditions remained the same, but the 4-digit group improved from worst to best of the 3 conditions. Analysis of Block 1 data indicated no differences among groups early in practice, $F(2, 57) < 1.0$. By Block 5, precision levels differed, $F(2, 57) = 4.05$, $p < .025$. Further analysis of the Block 5 data revealed a linear trend across precision levels, $F(1, 57) = 7.56$, $p < .01$. This linear trend contrasted with the curvilinear trend found in Block 5 of Experiment I. This disappearance of an optimum in Experiment II was consistent with the theories of both Adams (1971) and Bilodeau (1966, 1969) as described above.

The data for the 1-, 2-, and 4-digit conditions from Block 5 of both studies were combined for further analysis. A feedback precision effect, $F(2, 114) = 3.21$, $p < .05$, and an experiment effect, $F(1, 114) = 12.10$, $p < .01$, were found. An Experiment × Feedback Precision interaction was also shown, $F(2, 114) = 5.34$, $p < .01$. That interaction supported the suggested dependency between precision of feedback and postfeedback interval length. On the whole, these findings were consistent with an information-processing approach to the learning of a simple positioning task. Such an approach might further suggest that other variables could be expected to interact with feedback precision and/or duration of postfeedback interval. Information-processing load upon *S* during that interval might vary with response complexity, number of response alternatives, feedback transformations, duration or rate of presentation of feedback, idiosyncratic information-processing strategies, and related variables which are extensively reviewed by both Adams and Bilodeau.

TABLE 2

MEAN RESPONSE ERROR (IN IN.) FOR FEEDBACK CONDITIONS OVER TRIAL BLOCKS: EXPERIMENT II

Digits of feedback	Trial block 1	2	3	4	5
1	5.75	2.79	1.91	1.33	1.22
2	5.27	2.93	1.49	.88	.83
4	4.68	2.00	.88	.68	.54

Experiment III

The phenomenon of a feedback precision optimum found in Experiment I might have been idiosyncratic to the micrometer positioning task. A third study was then conducted with a different apparatus and procedure. An effort was also made to apply a more specific information-processing approach to the present problem. Bernstein, Blake, and Hughes (1968) view

information processing as occurring in 3 successive stages: filtering or detection, processing or stimulus analysis, and response selection. Separate mechanisms are postulated to operate at each stage and rarely operate in parallel (Decker & Rogers, 1973). The most likely locus for operation of feedback precision effects would be in the stimulus analysis and/or response selection stages during the postfeedback interval. By reducing the number of response alternatives to 1, while continuing to present extrinsic feedback at suprathreshold levels, varying precision levels might be studied relative to the stimulus analysis stage alone. A delayed single-choice reaction time task was then chosen for Experiment III. Occurrence of a precision optimum with such a task would then lend limited generality to the phenomenon and associate it in part with the theorized stimulus analysis stage of Bernstein et al.

Method

Subjects. Forty-five undergraduate volunteers from the University of Arizona were randomly assigned to 3 groups without regard to sex. Each was paid for his participation.

Apparatus. A modified Lafayette reaction time apparatus (Model 322B) was interfaced with an electronic timer (Hewlett-Packard Model 522B). Onset of the single blue stimulus light activated the timer. Depression of a telegraph key terminated elapsed time measurement and provided offset of the light. The calibrated timer measured elapsed time to the nearest .00001 sec.

Procedure. The *S*s were seated before a table with the index finger of their preferred hand resting upon a telegraph key. They were instructed that their task would be to learn to wait, following onset of the light, for a specific interval prior to depressing the key. Following a response, feedback was presented for direction (too soon or too late) and magnitude (in seconds) of timing error. Target time was set at 9 sec. ± .5 sec., but *S*s were not so informed. Ten trials were provided. The event sequence consisted of a verbal ready signal followed in 2 sec. by onset of the stimulus light, followed in turn by *S*'s response, feedback 36 sec. after the ready signal, and the next ready signal 5 sec. after termination of feedback. Feedback duration was about 4 sec., providing for a total trial cycle of 45 sec. When *S* responded during the target interval, he was told that he was correct and should repeat his performance on the next trial.

Design. Feedback grain was varied among the 3 groups. The 1-digit group received error rounded

TABLE 3

MEAN TIMING ERROR (IN SEC.) BY FEEDBACK CONDITION AND TRIAL BLOCK: EXPERIMENT III

Digits of feedback	Trial block				
	1	2	3	4	5
1	7.16	4.87	2.63	1.49	1.54
4	5.89	1.91	1.24	1.05	.89
8	7.24	2.87	2.12	1.77	1.45

to the nearest whole second. The 4-digit group was given feedback rounded to the nearest .001 sec., while 8-digit *S*s received feedback refined to the nearest .0000001 sec. In the latter group, the last 2 feedback digits were taken from a random numbers table because the timer was only accurate to 6 digits.

Results and Discussion

Table 3 presents mean absolute response error for each feedback condition over 5 2-trial blocks. No differences among groups were found in Block 1, F (2, 42) < 1.0. Further analysis indicated a practice effect, F (4, 168) $= 41.75$, $p < .01$, and a Practice × Feedback Precision interaction, F (8, 168) $= 2.43$, $p < .05$. No simple precision effect was found, F (2, 42) $= 2.07$. Examination of Block 5 data showed a curvilinear trend over precision levels, F (1, 42) $= 10.0$, $p < .01$, as was found in Experiment I. Optimal feedback precision was the 4-digit condition. Maximum precision produced intermediate level performance with the 1-digit group ranking lowest. The feedback precision effect found in Experiment I with the positioning task was then replicated for the timing task. The optima for the 2 task–procedure combinations were different despite a 7-sec. postfeedback interval for both studies. The higher 4-digit optimum obtained in Experiment III may have resulted from any number of differing variables between the 2 studies, including the shorter trial cycle and simplified response demand characteristics in the timing task. The deleterious acquisition effect of excessively precise feedback in the third study was less severe than in the first, perhaps for similar reasons. On the whole, Experiment III findings lend limited

generality to the feedback optimality phenomenon. The existence of such an optimum, in a task where extrinsic feedback was at suprathreshold levels and response selection demands were low, suggests that the information-processing mechanism affected may operate in part during the stimulus analysis stage described by Bernstein et al. (1968).

Further extension of the Bernstein et al. approach to simple skills tasks would suggest that the detection stage operates roughly between onset and offset of feedback. This feedback duration interval occurs between pre- and postfeedback intervals but has rarely been extracted or examined in simple positioning tasks. Previously established signal detection variables would presumably operate during this feedback presentation interval. Occurring next, during the early part of the postfeedback interval, would be the stimulus analysis stage. Its minimal length would be determined by feedback characteristics, long- and short-term verbal and motor memory demands, and possibly cognitive or problem-solving requirements in some tasks. Bernstein's response selection stage would occur last in the postfeedback interval, provided such time were available. The Bernstein et al. formulations may represent a convenient vehicle for deriving increasingly specific hypotheses as to the nature, duration, and sequencing of postfeedback interval events in simple motor-skills tasks. Such a microscopic analysis of within-trial events follows directly from the suggestions of bo[illegible] Adams (1971) and Bilodeau (1966, 1969) th[illegible] simple motor tasks have strong verbal-[illegible]nitive components active during the p[illegible]tfeedback interval.

REFERENCES

ADAMS, J. A. A closed-loop theory of m[illegible]arning. *Journal of Motor Behavior*, 1971, [illegible] 149.

BERNSTEIN, I. H., BLAKE, R. R., & HUGH[illegible] M. H. The effects of time and event uncer[illegible] on sequential information processing. *Pe[illegible]n & Psychophysics*, 1968, 3, 177–184.

BILODEAU, E. A., & ROSENBACH, J. H. *A[illegible]tion of response proficiency as a function o[illegible]ding error in informative feedback.* (Res. B[illegible] No. 53-21) Lackland Air Force Base, Tex[illegible] S. Air Force Human Resources Research Center, 1953.

BILODEAU, I. McD. Information feedback In E. A. Bilodeau (Ed.), *Acquisition of skill.* [illegible]ew York: Academic Press, 1966.

BILODEAU, I. McD. Information feedbac[illegible] In E. A. Bilodeau & I. McD. Bilodeau (Eds.), *[illegible]nciples of skill acquisition.* New York: Academic Press, 1969.

BOURNE, L. E., JR., & BUNDERSON, C. V. Effects of delay of informative feedback and length of postfeedback interval on concept identification. *Journal of Experimental Psychology*, 1963, 65, 1–5.

DECKER, L. R., & ROGERS, C. A., JR. Forced guidance and distribution of practice in sequential information processing. *Perceptual and Motor Skills*, 1973, 36, 415–419.

TROWBRIDGE, M. H., & CASON, H. An experimental study of Thorndike's theory of learning. *Journal of General Psychology*, 1932, 7, 245–258.

WEINBERG, D. R., GUY, D. E., & TUPPER, R. W. Supplementary report: Variation of postfeedback interval in simple motor learning. *Journal of Experimental Psychology*, 1964, 67, 98–99.

(Received June 25, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 609-614

PRODUCTION, ESTIMATION, AND REPRODUCTION OF TIME INTERVALS DURING INHALATION OF A GENERAL ANESTHETIC IN MAN[1]

NILLY ADAM,[2] ANGELINA D. CASTRO, AND DONALD L. CLARK

School of Medicine, University of Pennsylvania

The internal clock model for time judgments of short intervals describes adequately the relations among the methods of estimation, production, and reproduction observed under normal conditions. Since general anesthetics affect time productions by slowing of this hypothesized clock, this study investigated whether the relations predicted by the model would hold during inhalation of a general anesthetic—fluroxene. Six *S*s estimated, produced, and reproduced 3-40-sec. intervals before and during fluroxene inhalation. The model predicted the results from the majority of *S*s, thus supporting the internal clock concept. The anesthetic increased the slope of the production function and accordingly decreased the slope of estimation, leaving the slope of reproduction unchanged, thus maintaining the predicted relations among the methods. An alternative model, viewing time judgments as depending on memory content, could not account for the effects of the anesthetic on time reproductions seen with some *S*s.

The 3 most common measurement methods of time perception are estimation, production, and reproduction. The relationships among time judgments obtained through these 3 methods have been the subject of numerous investigations (for reviews, see Adam, 1971; Bindra & Waksberg, 1956). The most explicit formulation of these relationships was put forward by Carlson and Feinberg (1968). They based their reasoning on the assumption of a "clock" underlying time perception. Presumably, *S* uses the same clock to judge time intervals measured by these methods.

In estimation, the slope of the function depicting time estimates against objective time gives the ratio of the subjective to the objective clock. That is, in the function $T_E = k_E t_o$, where T_E = time estimates, t_o = objective time, and k_E = the slope or the ratio of the number of "subjective" seconds per each "objective" second.[3] Thus, if *S* receives a 2-sec. stimulus which he estimates as being 1 sec. long, his subjective/objective ratio is .5 and accordingly the obtained function $T_E(t_o)$ should subtend 26°30′ (arctan .5).

In production, *S* is expected to keep the same subjective/objective ratio as in estimation, although this is not necessarily so. If the same ratio is used, then when the function of time production—$T_P = k_P t_s$, where T_P = time produced in actual seconds, k_P = the slope, and t_s = the stimulus stated verbally—is plotted on the same graph as time estimates, the *y* axis now gives the objective seconds as produced by *S* and the *x* axis now gives the subjective seconds. Hence,

$$k_P = 1/k_E. \qquad [1]$$

Reproduction also depends on the same clock. One way of formulating the events during reproduction is assuming that first *S* estimates the presented interval and then produces the estimated interval. If this describes adequately the events in reproduction, then based on the previously

[1] This research was supported by U.S. Public Health Service Research Center Grant 5-P01-GM-15430-05. The authors appreciate the technical assistance of K. Van Horn and E. Dubroff.

[2] Requests for reprints should be sent to Nilly Adam, P.O. Box 4563, Philadelphia, Pennsylvania 19131.

[3] A linear function was chosen since time judgments at the stimulus range of the present study (3-40 sec.) have been shown by numerous investigators to follow a linear trend. This is also true for judgments obtained during anesthetic inhalation (Adam, Rosner, Hosick, & Clark, 1971). Intercepts are usually very small (Adam et al., 1971; Carlson & Feinberg, 1968) and are negligible for the purpose of the analysis.

stated relations between estimation and production, the following should be true:

$$T_R = k_R t_r, \qquad [2]$$

where T_R = the reproduced interval, t_r = the presented stimulus, and k_R = the slope of the function. Since first *S* is assumed to estimate the stimulus, this estimate is

$$T_E = k_E t_r. \qquad [3]$$

This estimate, T_E, in turn, is the stimulus for production; thus,

$$T_P = k_P T_E. \qquad [4]$$

Since the final value of reproduction is equal to the production of the estimated interval, it follows from Equations 2–4 that

$$T_R = T_P = k_P k_E t_r. \qquad [5]$$

It follows from Equations 2 and 5 that

$$k_R = k_P \cdot k_E. \qquad [6]$$

Since k_P has already been defined as $k_P = 1/k_E$ for the case where *S* uses the same clock rate in estimation and production, then

$$k_R = 1. \qquad [7]$$

In other words, the function of reproduction against time should always have a slope of 1.0 (i.e., 45°).

The relations delineated above have been shown to closely approximate the actual results obtained from normals and schizophrenics (Carlson & Feinberg, 1968). In a more recent paper, Carlson and Feinberg (1970) have shown that production and estimation may involve different "clock rates" when not obtained from the same *S*. Also, the slopes of production and estimation may vary with experience even when obtained from the same *S*. However, it was demonstrated again that if all 3 measures are obtained from the same individual the relations outlined above are maintained.

The purpose of this study was to use anesthetic drugs in order to observe whether the relations among the 3 methods still hold under conditions where it is known that at least one of these methods gives an altered slope. Adam, Rosner, Hosick, and Clark (1971) have shown that some general anesthetics increase the slope of production, that is, slow down the supposed clock.[4] The present experiment employed low concentrations of fluroxene —a general anesthetic. Time judgments were obtained before and during drug inhalation.

Method

Subjects. The *S*s were 6 young paid male volunteers participating in a study of the neurophysiological effects of fluroxene in conscious man.[5] After giving informed consent to participate, each *S* underwent a complete physical examination, blood count, urine analysis, chest X ray, electrocardiogram, and determinations of blood urea nitrogen, and serum glutamic oxaloacetic transaminase. Only *S*s within the normal range on each of these tests were accepted for study.

The *S* avoided all solids, liquids, and drugs for 12 hr. prior to the study. During the study, respiratory rate and end-tidal carbon dioxide were monitored. Blood pressure, heart rate, and rectal temperature were recorded throughout the study.[6] The anesthetic was delivered through a nonrebreathing circuit vented to the out-of-doors. A mouthpiece was used rather than a mask for easier control of drug concentrations. Dose was adjusted individually to allow testing. The *S* was tested when equilibrium was reached (end-tidal concentration within .1–.2% of the inspired concentration).

Material and procedure. Reproduction was tested by having *S* listen to a tone for a specified number of seconds. The *E* then started a second tone and *S* had to stop the tone by pressing a microswitch when he felt that the second tone equaled the first one in duration. For production, *S* activated a tone by pressing a microswitch which he released when he felt the duration of the tone was equal to an interval stated verbally by *E*. In estimation, *S* heard a tone activated by *E*. He then had to state its duration in seconds.

Reproduction was tested for 3-, 10-, 20-, and 40-sec. intervals, estimation for 3-, 20-, and 40-sec.

[4] This effect cannot be attributed to the slowdown which occurs with practice (Carlson & Feinberg, 1970). The removal of the anesthetic brought about a decreased slope, whereas the effect of practice on production alone would be to continuously increase the slope.

[5] D. L. Clark, A. D. Castro, and N. Adam. Neurophysiological effects of fluroxene. Manuscript in preparation.

[6] No correlation was found between these variables and time judgments in another study with other general anesthetics (Adam et al., 1971). Hence this aspect was not assessed in the present study.

intervals, and production for 6-, 20-, and 40-sec. intervals. Each S was tested with all 3 methods. The order of methods was counterbalanced over Ss. Stimuli for each method were presented randomly, with the constraint that not more than 2 identical stimuli were given consecutively.

Control measures were obtained first, with the anesthesia system hooked to S and delivering zero concentration. Testing during drug inhalation was done when equilibrium was achieved. The only time S gave a verbal response was during estimation. Usually it was possible to understand his answers despite the mouthpiece. If not, a throat microphone was used. If this was not effective, S was asked to signal his answer with his fingers.

When testing was completed, the mouthpiece was removed and S was taken to the hospital recovery room until ready for discharge.

Results

Time judgments were plotted separately for each individual for the 3 methods. The best-fitting linear functions for individual Ss for control and anesthetic conditions are shown in Figure 1. Table 1 gives the arctan of the slope of each function, the expected arctan slope of production according to the relation Slope Production = 1/Slope Estimation, and the expected arctan slope of reproduction according to the relation Slope Reproduction = Slope Estimation × Slope Production (see Equations 1 and 6). Intercepts were very small, with $M = .21$ and $SD = .77$. Since the anesthetic did not affect the intercept significantly (see also Adam et al., 1971), the intercepts are not shown in the table.

As seen, 4 Ss (RS, RO, KC, and RM) showed the expected relations among the 3 methods in both control and drug conditions. The deviations of the actual arctan slope of production from the expected values were, respectively, 2°10′, 2°10′, 4°20′, and 3°20′ for control, and 0°50′, 2°20′, 1°, and 1°10′ for the drug condition. The deviations of the arctan slope of reproduction from the expected values for these 4 Ss were 1°30′, 0°30′, 0°40′, and 0°10′ for control, and 0°20′, 3°50′, 3°20′, and 2°20′ for the drug condition. All of these values represent deviations of less than 10% from the expected values. There was no tendency for the actual values to be either consistently larger or smaller than the expected values.

Although the anesthetic did not alter the relations among the results from the 3 methods in these 4 Ss, it did affect the slopes of estimation and production. The arctan slope of production was increased in all 4 Ss, ranging from an increase of 1°30′ for RS to 13°50′ for KC. This confirms the results obtained by Adam et al. (1971) for the production method with several other inhalation anesthetics. The slope of estimation was decreased accordingly, ranging from a decrease of 3°20′ for RM to a decrease of 17°15′ for KC.

The 2 remaining Ss, however, deviated in some respects from the expected trends. For both TB and AS, all 3 control response curves had slopes smaller than 1.0 (arctan smaller than 45°) (see Figure 1). This is in contrast to the expected curves, where reproduction should subtend 45° and production and estimation should be symmetrical around the reproduction curve. This implies that both these Ss did not follow the relation shown in Equation 1. Table 1 gives the expected and obtained values for k_P.

Note that if $k_P = 1/k_E$, then the other relation derived from the model is $k_R = k_P \cdot k_E = 1$ (Equations 6 and 7). However, if $k_P \neq 1/k_E$ but if reproduction still follows the suggested sequence of estimation and production, then the relation $k_R = k_P \cdot k_E$ still holds, only it is not equal to 1 (Equation 6). The control results of TB and AS seem to fit this latter relation. The deviation of the obtained from the expected arctan k_R was 0°55′ for TB and 2°50′ for AS, a deviation of less than 10%.

The anesthetic affected the results in the predicted direction for production and estimation. The slope of production increased and that of estimation decreased. The only S to be tested at 2 different anesthetic concentrations was AS. As predicted, as concentration increased from 1.01% end-tidal to 1.21% end-tidal, the arctan slope of production increased from 49° to 63°. Accordingly, that of estimation decreased from 36°45′ to 31°25′, indicating the dose-

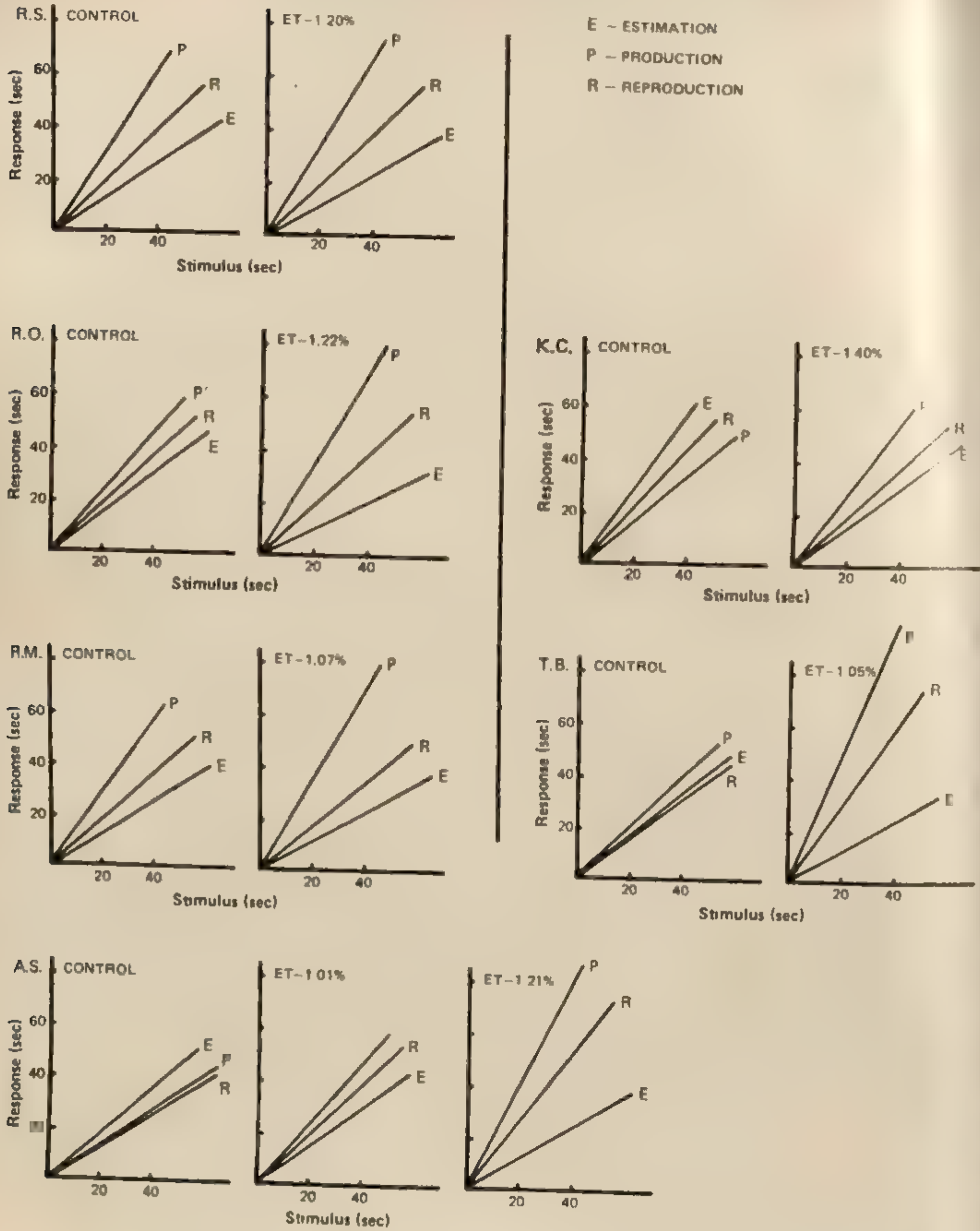


FIGURE 1. Best-fitting linear functions for estimation, production, and reproduction before and during fluroxene inhalation. (Data presented for individual *Ss*. The corresponding arctan slopes are given as obtained values in Table 1.)

related effect of the drug. The relation $k_R = k_P \cdot k_E$ was held also during the anesthetic condition for both *Ss*. The deviation of obtained from expected values was 1°20′ for TB and 3°10′ and 2°30′ for AS for the lower and higher drug concentration, respectively. Note that for AS at end-tidal concentration of 1.01% arctan k_R came very close to the expected 45° (obtained value = 43°50′). However, this

TABLE 1

ARCTAN SLOPE OF RESPONSE CURVES FOR INDIVIDUAL *S*s BEFORE AND DURING FLUROXENE INHALATION

S and condition	Order[a]	Arctan slope (in degrees): Obtained values: Estimation	Obtained values: Production	Obtained values: Reproduction	Expected values: Production[b]	Expected values: Reproduction[c]	Anesthetic concentration (%): Inspired	Anesthetic concentration (%): End-tidal
RS Control	R, E, P	34°30′	57°40′	44°30′	55°30′	46°		
Drug		30°	59°10′	44°20′	60°	44°	1.35	1.20
RO Control	P, R, E	38°20′	49°30′	43°30′	51°40′	43°		
Drug		27°20′	60°20′	44°10′	62°40′	40°20′	1.45	1.22
KC Control	P, E, R	54°15′	40°10′	47°30′	35°50′	48°10′		
Drug		37°	54°	43°	53°	46°20′	1.78	1.40
RM Control	E, P, R	32°30′	54°10′	41°40′	57°30′	41°30′		
Drug		29°05′	59°50′	40°10′	61	42°30′	1.37	1.07
TB Control	E, R, P	38°20′	43°40′	36°05′	51°42′	37°		
Drug		29°30′	67°40′	55°20′	60°32′	54°	—	1.05
AS Control	R, P, E	41°10′	33°10′	32°30′	48°50′	29°40′		
Drug		36°45′	49°	43°50′	53°14′	40°40′	1.11	1.01
		31°25′	63°	52°40′	58°34′	50°10′	1.39	1.21

[a] Abbreviations for order of presentation of methods: E = estimation, P = production, R = reproduction.
[b] Predicted on the basis of the relation Slope Production = 1/Slope Estimation.
[c] Predicted on the basis of the relation Slope Reproduction = Slope Estimation × Slope Production.

has no significance since further increase of concentration again removed k_R from 45° (see Table 1).

The order in which *S* had to perform the 3 methods had no effect on the results. Table 1 gives the order. Note that each *S* was tested in a different order. However, any order effect would be mainly due to the order of estimation and production in relation to one another. Note that the 2 exceptional *S*s, TB and AS, performed estimation and production in a different order.

DISCUSSION

Control results showed that the model first suggested by Carlson and Feinberg (1968) and delineated mathematically in the introduction, describes adequately the time behavior of 4 *S*s but only partially the results from the other 2 *S*s (TB and AS). For the first 4 *S*s all predicted relations were found. The obtained slope of production was reasonably close tlo the expected 1/Slope Estimation and the sope of reproduction was reasonably, close to the product of the slopes of estimation and production. This is in line with the findings of Carlson and Feinberg (1968, 1970) for the case when each *S* gave time judgments on all 3 methods. This gives support to the theoretical notion of a single "clock" underlying estimation and production and explaining reproduction as involving estimation of a presented interval followed by production of the estimated interval. The fact that in these 4 *S*s the anesthetic altered the slopes of production and estimation but without changing the expected relations among the 3 methods validated the results obtained under normal conditions.

Two *S*s, however, gave somewhat different results. In both cases, all 3 psychophysical methods had a slope smaller than 1.0. This implies that the slope of production did not conform to the predicted 1/Slope Estimation. This may mean one of two things. First, for some reason these 2 exceptional *S*s may have used a different clock rate for production

than for estimation. This, in turn, resulted in a slope for production that is not the inverse of that for estimation. Another possibility is that the same clock was used for both methods but some unidentified response or processing variables affected the final result. Such variables could, of course, affect either estimation or production or both. It is reasonable to consider estimation as a process involving "starting the clock" and obtaining a reading when the stimulus is over; production then may be conceptualized as a process in which S produces his own imaginary stimulus which he estimates every so often and compares the result with the required interval. He stops when the 2 match. If so, then obviously different response variables may affect the 2 processes. The present results do not give any clue as to why only 2 out of 6 *Ss* were affected in this way.

Since for these 2 *Ss* k_P did not equal $1/k_E$, it follows from Equation 6 that $k_R = k_P \cdot k_E$ but need not equal 1.0. This relation was found to describe adequately the results from the 2 exceptional *Ss*, thus supporting the notion that reproduction depends on estimation with later production of the estimated interval. The anesthetic increased the slope of production and decreased that of estimation, as it did with the other 4 *Ss*.

No correlation was found between the serial order of response and the magnitude of a time judgment. Also, the order of presentation of methods did not seem to have a noticeable effect. It is particularly important that the order of estimation and production was different for the exceptional *Ss*.

Another approach to the mechanism underlying time perception suggests that time judgment depends on memory content (Frankenhaeuser, 1959). Thus, the obtained linear functions describing time judgments of periods longer than the immediate present reflect the decay of memory traces. According to such a concept, the anesthetic altered the slopes of estimation and production because the memory content deteriorates faster under the anesthetic. However, Frankenhaeuser has not presented any specific model for predicting the relationships among the 3 psychophysical methods, nor is it clear from her work how reproduction is carried out. If she adhered to the reproduction process as presented in the present article, her model would predict the same anesthetic effects on reproduction as the clock model. How[illegible] there is an alternative way to conceptuali[illegible] reproduction which fits her framework. [illegible]m the point of view of memory theory, [illegible]luction may be thought of as reflectin[illegible]gnition processes; that is, the second [illegible]is is found matching to the first stimul[illegible] when it is recognized as being similar to the first stimulus already stored in mem[illegible] This latter possibility fits well the data [illegible]4 of our *Ss*. As part of another study, [illegible] were also tested for performance on a [illegible]gnition task of 6-word lists during fluroxene inhalation. The effects were negligible for all *Ss*. Thus, the fact that the slope of reproduction did not change in the anesthetic condition in 4 *Ss* can be accounted for by the negligible effects of fluroxene on recognition at the studied concentrations. However, 2 of our *Ss* did show altered reproduction slopes in the anesthetic condition. These changes were predictable on the basis of the relations among the 3 psychophysical methods provided by the clock model. These results cannot be explained on the basis of impaired recognition performance since the latter was hardly affected. Some anesthetics do have a marked effect on recognition even at very low concentrations (Adam, 1973). It should be interesting to observe behavior on time judgments during inhalation of these anesthetics.

REFERENCES

ADAM, N. Mechanisms of time perception. *T.-I.-T. Journal of Life Sciences*, 1971, **1**, 41–52.

ADAM, N. Effects of general anesthetics on memory functions in man. *Journal of Comparative and Physiological Psychology*, 1973, **83**, 295–306.

ADAM, N., ROSNER, B. S., HOSICK, E. C., & CLARK, D. L. Effects of anesthetic drugs on time perception and alpha rhythm. *Perception & Psychophysics*, 1971, **10**, 133–136.

BINDRA, D., & WAKSBERG, H. Methods and terminology in studies of time perception. *Psychological Bulletin*, 1956, **53**, 155–159.

CARLSON, V. R., & FEINBERG, I. Individual variations in time judgment and the concept of an internal clock. *Journal of Experimental Psychology*, 1968, **77**, 631–640.

CARLSON, V. R., & FEINBERG, I. Time judgment as a function of method, practice, and sex. *Journal of Experimental Psychology*, 1970, **85**, 171–180.

FRANKENHAEUSER, M. *Estimation of time.* Stockholm: Almquist & Wiksell, 1959.

(Received June 26, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 615–620

MULTIDIMENSIONAL ANALYSIS OF CHOICE REACTION TIME JUDGMENTS ON PAIRS OF ENGLISH FRICATIVES

FREDERICK F. WEINER [1]
Department of Speech and Hearing Science, University of Maryland

SADANAND SINGH
Ohio University

Choice reaction time (CRT) judgments of *same–different* responses were measured for all possible pairs of nine English consonants by two groups of 10 *S*s each. A multidimensional analysis (IND SCAL) of the CRT values, for the first group of *S*s showed that the magnitude of *S*s' CRTs was governed by the distinctive feature differences between consonant pairs. A four-dimensional analysis revealed that the distinctive features, *voicing*, *sibilant*, *front/back* and *palatal* were those used by the group. Examination of the individual *S*s showed differing patterns of weighting these features. Analysis of the second group of 10 *S*s served to cross-validate the results obtained from the first group. Results showed no significant difference in the relationship between CRT and distinctive feature differences for the two groups. Comparison between these results and the 1969 results of Chananie and Tikofsky seems to suggest that CRT is a good measure of phoneme similarity when extraneous factors contributing to CRT are minimized.

Judgments of consonant similarity have been used to find a set of phonological features that may be distinctive within a listener's repertoire (Singh, Woods, & Becker, 1972). These judgments of similarity have been made on the basis of subjective evaluations using methods of magnitude estimation and psychological scaling (Singh et al., 1972) or by obtaining perceptual errors of direct judgments while listening to signals acoustically distorted by noise or filtering (Miller & Nicely, 1955; Singh, 1966, 1971).

The results of several investigations have indicated that choice reaction time (CRT) varies directly with the similarity of stimulus alternatives, (Crossman, 1955; Thurmond & Alluisi, 1963; Weiner, 1973). Reaction times are short when discriminating between items which are obviously dissimilar and increase when discriminating between similar stimuli.

Chananie and Tikofsky (1969) attempted to evaluate the similarity of consonants by relating CRT with the number of distinctive feature differences between consonants. They were unable to show that CRT was a measure of interphonemic similarity determined by distinctive feature differences.

[1] Requests for reprints should be sent to Frederick E. Weiner, Department of Speech and Hearing Science, University of Maryland, College Park, Maryland 20742.

The distinctive feature system used was that of Miller and Nicely (1955). Recently, investigators have demonstrated that this feature system is a relatively poor predictor of phonemic perception (Singh & Becker, 1972; Singh et al., 1972; Wickelgren, 1966). Others have varied the experimental conditions and found this system a good or better predictor of phoneme perception than other feature systems (Singh, 1970; Wang & Bilger, 1973).

Rather than beginning with an a priori feature system, such as Miller and Nicely's (1955), to evaluate similarity, the present investigation was designed to find, by means of multidimensional scaling technique, the feature system used by a group of *S*s and then to determine whether CRT relates to similarity in terms of the features retrieved.

Method

Subjects. Two groups of 10 undergraduate native American English speakers without history of hearing loss served as *S*s. The responses of the first group of *S*s served as an input to the multidimensional analysis for purposes of retrieving the group's perceptual feature system. The responses of the second group of *S*s were used to cross-validate the relation between CRT and consonantal similarity observed in the first group of *S*s.

Stimuli. The stimuli were pairs of consonant–vowels (CVs) composed of one of the following English fricatives: /f,v,θ,ð,s,z,ʃ,ʒ,h/ followed by the vowel /ɑ/. Half of the stimuli were constructed so

that both pairs of CVs for each item were identical. The remaining items were constructed so that the fricative of the second member of each pair was different from the fricative for the first member of that pair. Thus, 72 stimuli were recorded on channel 1 of a two-channel tape recorder. A pure tone was recorded on channel 2 at the initiation of the second member of each pair. All stimuli were presented so that the members of each pair were separated by 1 sec., with 5 sec. elapsing between stimuli.

Apparatus. The stimuli were presented by way of a Sony (TC 200) stereo tape recorder with a frequency response of 50–10,000 Hz. ± 3 db. and an overall frequency response of 50–14,000 Hz. The *S*s heard the stimuli through TDH-39 headphones. The output from channel 2 was connected to a pulse generator which in turn was connected to a (model DP 210) Hickock time interval meter. Each time a signal appeared on channel 2, a pulse was sent from the pulse generator to start the time interval meter. The *S*s were seated in front of a reaction time box having one button labeled "S" and the other labeled "D." The reaction time box was connected to the time interval meter so that depression of one of the two buttons stopped the meter. Choice reaction time, then, was measured from the initiation of the second CV to the eventual response of pushing one of the two buttons. Response times and button choice were recorded on a Hickock (PR 4900) printer connected to the time interval meter.

Subjects were instructed to respond by pressing the button marked "S" if the two phonemes they heard were identical or to press the button marked "D" if the two phonemes were not the same. They were told to respond as quickly as possible without committing errors. Each *S* was presented with 30

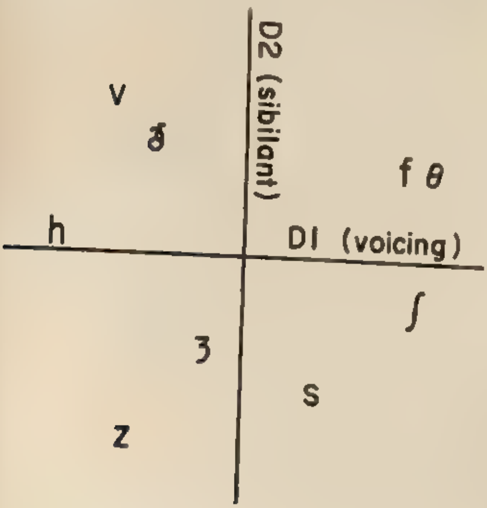


FIGURE 1. Plotting of Dimension 1 (D1) and D2 of the four-dimensional IND SCAL configurations showing *voicing* and *sibilant*.

practice items and was encouraged to respond below 500 msec. Most *S*s made their determinations before the termination of the second syllable. Thus, *S*s with unusually long CRTs were trained to make their decisions before completion of the second syllable for each stimulus. The *S*s then participated in the experiment, which consisted of four randomized presentations of each stimulus pair.

RESULTS AND DISCUSSION

The median of the four CRTs for each stimulus pair was taken to represent *S*'s score on that pair. Utilizing these scores, 10 symmetric matrices for each *S* in Group 1 were analyzed by Individual Scaling (IND SCAL) developed by Carroll and Chang (1970). This procedure is a multidimensional scaling program designed to determine the number of dimensions *S*s utilize in making psychological judgments. The output provides stimulus and *S* weightings on each of the dimensions of an *n*-dimensional space.

The choice of dimensionality was based on the criteria of uniqueness and interpretability of solution. A unique solution is one in which the stimuli and *S* points in a given dimensional space converge to the same solution irrespective of the random starting configuration. Interpretability is determined by the solution yielding the most definable phonetic features. In the present investigation a four-dimensional solution was considered unique where Dimension 1 (D1) was interpreted as *voicing*; Dimension 2 (D2) as *sibilant*; Dimension 3 (D3) as *front back*; and Dimension 4 (D4) as *palatal*.

Figure 1 is a plotting of D1 versus D2. From this representation it can be seen that the voiceless consonants of the set /f,θ,s,ʃ/ are separated from the voiced consonants /v,ð,z,ʒ,h/. Although /h/ is usually considered voiceless in English, perceptually it was positioned at the extreme end of the voicing dimension. The most probable explanation for the placement of /h/ is that /h/ is homorganic with /ɑ/, with which /h/ was combined. The phoneme /h/ is also the only consonant which is considered minus-consonantal (Chomsky & Halle, 1968), a feature regularly assigned to vowels. Thus, *S* judgment may have

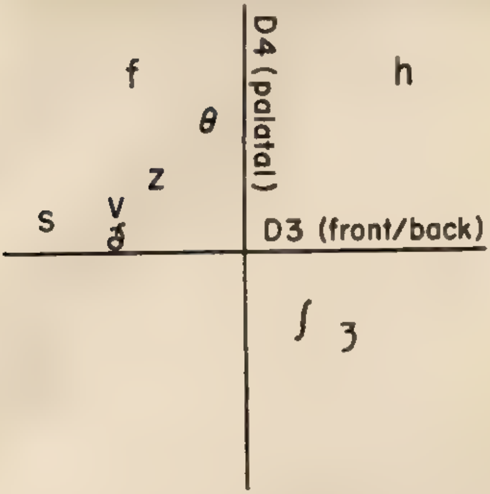


FIGURE 2. Plotting of Dimension 3 (D3) and D4 of the four-dimensional IND SCAL configurations showing *front/back* and *palatal*.

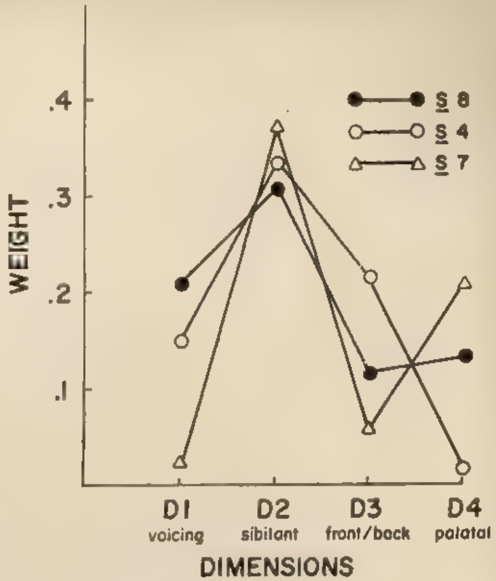


FIGURE 3. Plotting of individual *S* weightings as a function of distinctive features retrieved for *S*s 4, 7, and 8.

been grossly influenced by /ɑ/ each time /hɑ/ was compared with other CV sequences. Dimension 2 is clearly a *sibilant* dimension wherein the sibilants /s,z,ʃ,ʒ/ are separated from the nonsibilant fricatives /f,v,θ,ð,h/.

Figure 2 is a plotting of D3 versus D4. Dimension 3 can be labeled as a *front/back* dimension where the consonants /f,v,θ,ð,s,z/ produced anteriorly in the oral cavity are separated from the consonants /ʃ,ʒ,h/ which are produced posteriorly. Obviously, these *S*s were not using other distinctions of place of articulation on this dimension. For example, in Figure 2 the /θ/ is positioned closer to *back* than /s/. Hence, the extent of the psychological reality of articulatory features for these *S*s was limited only to a *front/back* distinction. On D4 the *palatal* consonants /ʃ,ʒ/ are separated from the nonpalatal consonants /f,v,θ,ð,s,z,h/. Thus more than one place dimension was utilized by these *S*s.

The four dimensional solution discussed above was found to account for 72% of the variance in *S* responses. Of this 72%, D1 (voicing) accounted for 24%; D2 (sibilant) 19%; D3 (front/back) 15%; and D4 (palatal) 14%. An examination of the individual *S* weights of these dimensions revealed three different trends. These three trends are reported in Figures 3, 4, and 5.

Figure 3 shows that *S*s 4, 7, and 8 weighted D2 the highest. Thus, incon-

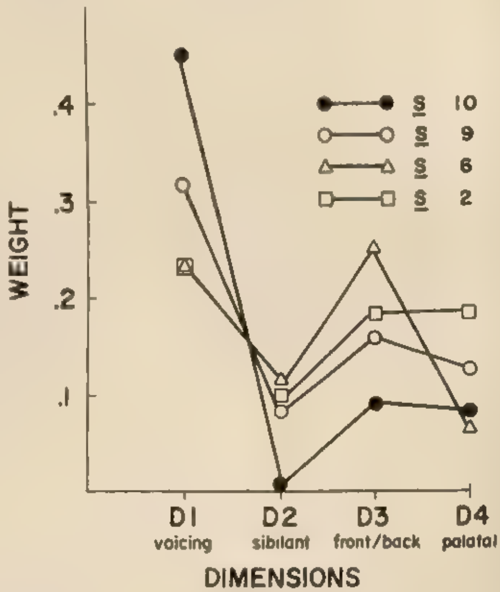


FIGURE 4. Plotting of individual *S* weightings as a function of distinctive features retrieved for *S*s 10, 9, 6, and 2.

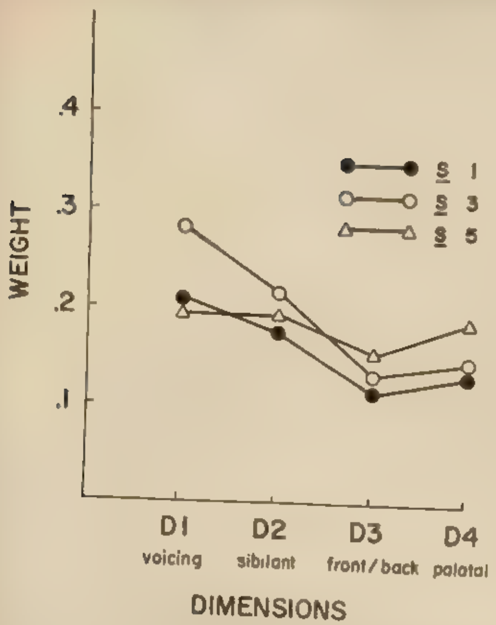


FIGURE 5. Plotting of individual *S* weightings as a function of distinctive features retrieved for *S*s 1, 3, and 5.

sistent with the overall trend, these *S*s were weighting *sibilant* inordinately high. In contrast, *S*s 2, 6, 9, and 10 in Figure 4 weighted *voicing* high and *sibilant* low. Again, in contrast with the overall trend, these four *S*s weight the *sibilant* dimension lowest. The weightings of the third group of *S*s are shown in Figure 5, where *S*s 1, 3, and 5 seem to confirm the overall trend of the general stimulus space.

In the past, various investigations have attempted to determine a feature system and the relative importance of the features for a group of *S*s (Miller & Nicely, 1955; Singh & Black, 1966; Wickelgren, 1966). Results of this study show that, while it is possible to determine a feature system for a group, it is difficult to draw conclusions about the relative weightings of features for that group. This difficulty arises from the evidence presented in Figures 3, 4, and 5 wherein it can be seen that group feature weights represent a mean of extreme individual weightings. Therefore, while it is interesting and possible to determine the feature system of a group, it is much more revealing to examine the way the individual *S*s weight these features. That is to say that examination of individual *S*s provides information regarding their perceptual strategy, which may be different from the strategy of the other *S*s.

Even though the weightings of the individual *S*s were different on each dimension, the general conclusion is that all *S*s utilized all four dimensions. On the basis of the four dimensions recovered in this study, it is assumed that CRT judgments can be used as a criterion for determining phonemic similarity. In a phonological system involving binary features, a pair of CV syllables, which is different by any one of the four features, is considered minimally distinct and thus should yield a larger CRT than a pair of CV syllables having a two-feature difference. This assumption was not validated by Chananie and Tikofsky (1969) who performed a similar study using the Miller and Nicely (1955) distinctive feature system. Figure 6 shows CRT as a function of "zero", one, two, three, and four feature differences. Zero difference represents the phoneme pairs /f–θ and

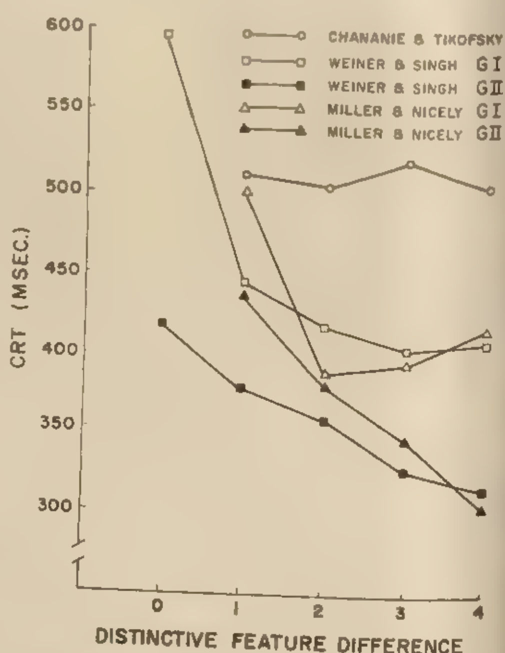


FIGURE 6. Choice reaction time as a function of distinctive features for data collected from groups one (GI) and two (GII) in the present study as compared with a previous study of Chananie and Tikofsky (1969).

/v–ð/ since these pairs remained indistinct on all of the four dimensions reported in this study. A one-feature difference represents a difference by any one of the four features. The two-, three-, and four-feature differences are combinations of any of the features.

The top curve in Figure 6, taken from Chananie and Tikofsky (1969), clearly shows a lack of relationship between CRT and distinctive feature differences. The curve labeled "Weiner & Singh GI" represents the mean CRTs for pairs of consonants representing zero, one, two, three, and four feature differences obtained for the first group of *S*s. A single-factor, repeated measures, analysis of variance was performed to test the difference between these mean CRT scores. The results showed that the mean differences were significant, F (4, 36) = 16.08, $p < .05$. A Tukey (*a*) (Winer, 1962) showed that CRT values for zero-feature differences were significantly greater than values associated with any other feature difference. In addition, the CRT value associated with one feature difference was significantly greater than differences of three and four features.

Since the above mentioned curve in Figure 6 was derived from the same data utilized to find the feature system, the slope was expected. It was therefore deemed necessary to cross-validate these results. The bottom curve is a plotting of the second group of *S*s. To determine whether these functions showed similar relationships, a comparison of slopes procedure (Dixon & Massey, 1969) was performed. The result was that the two beta coefficients were not different from each other, t (6) = .494, $p < .05$. Thus, the slopes were not significantly different from each other. There was a difference, however, in terms of the magnitude of the two curves. This is attributed to substituting photoelectric cells for the "push-type" response buttons used by the first group; a change made before realizing that a cross-validation would be necessary. However, for the purposes of this investigation, magnitude of CRT is not important. The purpose was to determine whether CRT was a function of distinctive feature differences.

A single-factor repeated measures analysis of variance was performed to test the difference between the mean values of CRT obtained for the second group. The results showed that the mean differences were significant, F (4, 36) = 11.17, $p < .05$. A Tukey (*a*) showed that CRT values for zero-feature differences were significantly greater than values associated with any other feature difference. Additionally, one-feature difference was significantly greater than a two-, three-, and four-feature difference; while differences of two, three, and four features were not different from each other.

Since a number of investigations have demonstrated that psychological judgment of the proximity of a phoneme pair is a function of the phonological difference between them, it is surprising that Chananie and Tikofsky (1969) did not find such a relationship. In an attempt to determine the reason for this discrepancy, we plotted the CRT values obtained from the two groups of *S*s in the present investigation as a function of the Miller and Nicely feature system as plotted by Chananie and Tikofsky. The curves labeled "Miller & Nicely GI" and "Miller & Nicely GII" in Figure 6 represent Miller and Nicely feature system plottings of the *S* data in the present investigation of groups one and two, respectively. Not only is it apparent from these results that the differences in features play a role in determining mean reaction times, but there is no significant difference between the slopes of Weiner & Singh GI and Miller & Nicely GI, t (5) = .524, $p < .05$, or Weiner & Singh GII and Miller & Nicely GII, t (5) = .621, $p < .05$. Thus it appears that either system of distinctive features explained the similarity of phonemes as evaluated by CRT. This finding is expected and should occur using any distinctive feature system since all feature systems draw on the same articulatory events.

The difference between the results of the present investigation and that of Chananie and Tikofsky (1969) may be attributed to their level of reaction time which was considerably higher than the reaction time obtained in the present study. Figure 6

suggests that the number of features distinguishing one phoneme from another becomes more important the lower the overall magnitude of reaction time. The magnitude was highest in the Chananie and Tikofsky study and number of features played no role. The magnitude was lower in Weiner & Singh GI and there was a slope. The magnitude was lowest in Weiner & Singh GII and the slope was steepest. In other words, it appears that when extraneous factors contributing to reaction time are minimized the differences in number of features begins to play a role.

REFERENCES

Carroll, J., & Chang, J. Analysis of individual differences in multidimensional scaling via an *n*-way generalization of Eckrt-Young decomposition. *Psychometrica*, 1970, **35**, 283–319.

Chananie, J., & Tikofsky, R. Choice response time and distinctive features in speech discrimination. *Journal of Experimental Psychology*, 1969, **81**, 161–163.

Chomsky, N., & Halle, M. *The sound pattern of English*. New York: Harper & Row, 1968.

Crossman, E. The measurement of discriminability. *Quarterly Journal of Experimental Psychology*, 1955, **7**, 176–195.

Dixon, W. J., & Massey, F. J. *Introduction to statistical analysis*. New York: McGraw-Hill, 1969.

Miller, G., & Nicely, P. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 338–352.

Singh, S. Crosslanguage study of perceptual confusion of plosive phonemes in two conditions of distortion. *Journal of the Acoustical Society of America*, 1966, **40**, 635–656.

Singh, S. Interrelationships of English consonants. In B. Hala, M. Romportl, & P. Janota (Eds.), *Proceedings of the sixth international congress of phonetic sciences*. Prague: Publishing House of the Czechoslavak Academy of Sciences, 1970.

Singh, S. Perceptual similarities of minimal phonemic differences. *Journal of Speech and Hearing Research*, 1971, **14**, 113–124.

Singh, S., & Becker, G. A. A comparison of four feature systems using data from three psychological methods. *Journal of Speech and Hearing Research*, 1972, **15**, 821–829.

Singh, S., & Black, J. Study of [illegible]6 intervocalic consonants as spoken and rec[illegible]ed by four language groups. *Journal of the A[illegible]stical Society of America*, 1966, **39**, 372–387.

Singh, S., Woods, D., & Becke[illegible] Perceptual structure of 22 prevocalic Engl[illegible] consonants. *Journal of the Acoustical Society of America*, 1972, **52**, 1698–1713.

Thurmond, J., & Allusi, F. Choice time as a function of dissimilarity and discriminability. *Canadian Journal of Psychology*, 1963, **17**, 326–337.

Wang, M., & Bilger, R. Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America*, 1973, **54**, 1248–1266.

Weiner, F. Choice reaction time as a function of frequency differences. *Perceptual and Motor Skills*, 1973, **36**, 307–310.

Wickelgren, W. Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, 1966, **39**, 388–398.

Winer, B. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received July 2, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 621–628

INTERACTIONS AND RANGE EFFECTS IN EXPERIMENTS ON PAIRS OF STRESSES:

MILD HEAT AND LOW-FREQUENCY NOISE [1]

E. C. POULTON [2] AND R. S. EDWARDS

Medical Research Council, Applied Psychology Unit, Cambridge, England

Twelve men performed 3 tasks in 102-db.(C) low-frequency noise, at 38°/33° C. (100°/92° F.), and with the two stresses combined, as well as in a control condition. The three tasks were tracking with peripheral lights, the five-choice task, and visual vigilance, in that order. The low-frequency noise had a beneficial effect upon all three tasks. It interacted with the mild heat on the tracking, and on false detections in the vigilance task. The results are related to behavioral arousal. When compared with a previous experiment on mild heat and loss of a night's sleep, performance in the control conditions was found to be influenced by the stresses included in the within-subjects experimental designs. This raises doubts about the validity of the interactions.

A number of experiments examine the interactions of pairs of stresses (Grether, 1970; Poulton, 1970, chap. 25). In some experiments reliable interactions are reported. In other experiments the effects of a pair of stresses appear to add. The combination of heat and noise does not produce reliable interactions, except possibly in the 1946 experiment by Viteles and Smith (see Wilkinson, 1969b, p. 267). The present experiment investigates the combination again, using tasks which are found to reveal reliable interactions between other pairs of stresses.

The left side of Figure 1 shows the results of a hypothetical experiment on the interaction of two stresses A and B. The control condition 00 is represented at the bottom. Stress A alone at A0, and stress B alone at OB, both increase the number of errors. Errors are rather more frequent still at AB, where the two stresses are combined. The interaction is reliable if the combined effect of A and B, represented by the height $(x + y + z)$, is reliably different from the sum of the two separate effects, represented, respectively, by the heights z and $(y + z)$. This reduces to x being reliably different from z. Since x represents the effect of A in the presence of B, and z represents the effect of A alone, an equivalent statement of the interaction is that the effect of A in the presence of B is reliably different from the effect of A alone.

Where reliable interactions are reported in the literature, the combined effect upon performance is always smaller than the sum of the two separate effects (Grether, 1970; Poulton, 1970). The only apparent exception is in experiments with knowledge of results as one of the two stresses. The relationship in Wilkinson's (1961) experiment on loss of sleep paired with knowledge of results, is illustrated on the right of Figure 1. In the first illustration on the right, A stands for the stress of no sleep

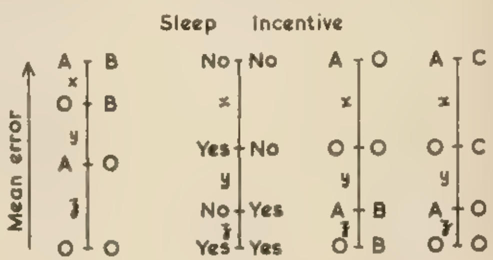


FIGURE 1. Interactions between two stresses.

[1] This work forms part of a program of research for the Royal Naval Personnel Research Committee. The experimental subjects were supplied by the British Royal Navy. The authors are grateful to A. Carpenter for building the noise equipment and for advice on its use. Muriel M. Woodhead kindly tested the hearing of three people with and without a thermistor in the external auditory meatus.

[2] Requests for reprints should be sent to E. C. Poulton, M.R.C. Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, England.

on the previous night, and B stands for the stress of immediate knowledge of results whenever the man makes an error, or does not respond for 1.5 sec. The loss of sleep increases errors, while the knowledge of results reduces errors. As usual the combined effect, represented here by $-y$, is smaller than the sum of the two separate effects, represented, respectively, by x and $(-y-z)$. This is equivalent to stating that zero is smaller than $(x-z)$, or that x is greater than z.

Wilkinson's (1961) alternative description of the interaction is illustrated on the extreme right of the figure. An O has been substituted for B, and a C for O. The C stands for loss of incentive, just as A stands for loss of sleep. Here the combined effect is represented by $(x+y+z)$. It is greater than the sum of the two separate effects, represented, respectively, by z and $(y+z)$, because x is greater than z.

Fortunately in most experiments on interactions between stresses, it is clear which conditions present the stresses. It is only with knowledge of results that there is any doubt. Probably this is because knowledge of results tends to be more effective when a person has been deprived of it, than it does when it has always been given (Brown, 1966). A within-subjects experimental design improves performance with knowledge of results, and depresses performance with no knowledge. Neither of the two changed levels of performance can be taken as the true baseline level of performance.

Range Effects

The observed heights of x and z may be distorted by a range effect. All experiments investigating the interactions between pairs of stresses use within-subjects designs. The same people perform in a control condition, under each stress singly, and under the combined stresses. Within-subjects designs can produce range effects which bias the results (Poulton, 1973). Performance in one condition is influenced by the other conditions included in the experimental design. A subsidiary aim of the present experiment is to look for possible range effects. Part of the experiment is designed to be identical with a previous experiment on mild heat and loss of sleep (Poulton, Edwards, & Colquhoun, in press). A reliable difference in the control condition which is common to the two experiments, would indicate the influence of range effects.

Low-Frequency Noise

Low-frequency or green noise is weighted heavily against the higher frequencies. It is used to reduce the chances of a permanent hearing loss, while maintaining a high overall sound-pressure level. Apparatus and speech sounds are not masked as effectively as they are by white noise, which has a flat spectrum. Low-frequency noise arouses the man, without isolating him from his auditory environment. Thus it may have a rather different effect upon performance from the conventional white noise. This is an additional point investigated here.

Method

Tracking with peripheral lights. Two of the three experimental tasks are described in detail in a previous paper (Poulton et al., in press). In the tracking task (Hockey, 1970a, 1970b) a target marker moves irregularly from side to side. In the present experiment the top track frequency is about 45 cycles per minute (cpm) instead of the 25 cpm used by Poulton et al. The man has to keep a response marker in line with a target marker, by raising and lowering a lever with his right hand. He is told that this is his main task. Time on target is noted for each 5-min. period.

The six peripheral lamps are arranged in a semicircle. Each lamp is surrounded by a lightproof case, except for a white circle of tape, 1.5 cm. in diameter, which lies in front of it. When the lamp is switched on, the white circle turns yellow. The brightness of the circle increases from 1 to 7 cd/m² (.3 to 2 equivalent ftc.). In the present experiment the man is provided with a chin rest. When he places his chin on the chin rest and looks at the central tracking display, the two nearest lamps lie 20° out from his line of sight, one on each side. Each lamp lights up on average once every 30 sec. The remaining four lamps, two on each side, lie, respectively, 50° and 80° out from the line of sight. Each lights up on average once every 2 min. The time interval between lamp signals varies irregularly from 6 to 24 sec. in steps of 3 sec. When a lamp lights up, the man has to extinguish it as quickly as possible by pressing the corresponding microswitch with a finger of his left hand. His response time is recorded. The combined task lasts 30 min.

The five-choice task. For this task (Leonard, 1959) five neon lamps are arranged in a pentagon of side 5 cm. They are mounted on an almost vertical panel. Five brass discs, each 3 cm. in diameter, are arranged in a corresponding pentagon with 9 cm. between centers. They are mounted flush in a horizontal panel which lies directly in front of the neon lamps. The man uses a stylus 18 cm. long to tap the disc corresponding to the neon lamp which is lit up. Tapping any disc extinguishes the lamp and lights up the next lamp in the series. The order of lamps in the series is irregular. The man has to work at the task without stopping for 30 min. The number of correct taps, and the number of times the wrong disc is tapped in error, are recorded automatically and noted for each 5-min. period. Gaps of 1.5 sec. or longer between responses are also recorded.

Wilkinson visual vigilance task. The task is a visual analogue of the Wilkinson (1969a) auditory vigilance task of the previous paper (Poulton et al., in press). A white circular window 1.5 cm. in diameter lies in the middle of a black panel 18 × 20 cm. The panel slopes backward at an angle of 45°. Immediately behind the window is a neon lamp The window has a brightness of 2 cd/m² (.5 ftL.) when the lamp is off. The window turns yellow and increases in brightness to 30 cd/m² (8 ftL.) when the lamp is on. The black panel has a brightness of .15 cd/m² (.04 ftL.). The neon lamp lights up for .4 sec. every 2 sec. A signal is a shorter flash of only .3 sec. There are five signals, scattered irregularly during each 7.5 min. The task lasts 30 min.

In the present experiment the man has three push buttons to indicate his degree of confidence, instead of the single push button of the previous experiment. When he detects a signal, he has to press one of the three buttons as quickly as possible. The three push buttons have diameters of 1.2 cm. They are mounted in a row 8 cm. below the circular window, with their centers separated by 4 cm. The blue push button on the right indicates certain signal. The green push button in the middle indicates probable signal. The yellow push button on the left indicates possible signal. A record is kept of each response button pressed. The response times of correct responses are also noted.

Experimental subjects and design. The 12 naval enlisted men have ages ranging from 18 to 28 yr., median age 21 yr. The experimental design is given in Table 1. There is a practice control condition before the start of the Latin square, as in the previous experiment (Poulton et al., in press). This enables the groups of naval enlisted men used in the two experiments to be compared directly, before they commence the conditions of the Latin square. The Latin square ensures that nobody works in noise on two successive days.

The hot and cool climates are similar to those of the previous paper (Poulton et al., in press). The hot climate has dry-bulb/wet-bulb temperatures of 38°/33° C. (100°/92° F.). This gives an Effective Temperature for men stripped to the waist of 34° C. (93° F.). The cool climate has temperatures of

TABLE 1

Experimental Design for Two-Week Period

Team[a]	Main practice (Wednesday)	Test 1 (Thursday)	Test 2 (Monday)	Test 3 (Tuesday)	Test 4 (Wednesday)
A	Control	Noise	Heat	Noise + Heat	Control
B	Control	Control	Noise + Heat	Heat	Noise
C	Control	Heat	Noise	Control	Noise + Heat
D	Control	Noise + Heat	Control	Noise	Heat

[a] In all cases, $n = 3$

20°/17° C. (68°/62° F.). This gives an Effective Temperature for men fully clothed of 19° C. (66° F.).

The low-frequency noise is presented from 2 loudspeakers, each 35 cm. in diameter. They are mounted respectively in the wall and ceiling of the hot room. The loudspeaker in the wall is at head height. It is on the man's right when he is tracking, and behind him when he rotates 90° to his left to perform the two remaining tasks. The ceiling loudspeaker is above his head, slightly to his left when he is tracking, slightly in front of him for the other 2 tasks.

The average intensity of the low-frequency noise is about 102 db. on the C scale of a Dawe 1400E sound-level meter. The intensity varies slightly, depending upon exactly where the man has his head. The noise is attenuated about 7.5 db. per octave by a lowpass filter. The noise had a rumbling quality. It is possible to converse by shouting at a distance of 1 m. In tracking with peripheral lights it is possible to hear the noises made by the pursuitmeter program, and to hear the clicks made by pressing the response switches for the lights. With the five-choice task it is possible to hear the difference between tapping a brass disc, and tapping the panel in which the discs are mounted.

The fans which circulate the air produce an average sound-pressure level of about 80 db. (C) in the quiet condition. Here the frequency spectrum is fairly flat up to 1 kHz. Above 1 kHz. the intensity is attenuated about 7.5 db. per octave.

Procedure. This follows the procedure of the previous experiment (Poulton et al., in press). Each man is always tested at the same time of day. One man of each team is tested in the morning, the other two men in the afternoon. The order of the three tasks is always the same: tracking with peripheral lights, five choice, and Wilkinson vigilance.

Thermistors are worn in both ears, to record the temperature of the external auditory meatus. Wearing a thermistor does not affect the threshold of hearing.

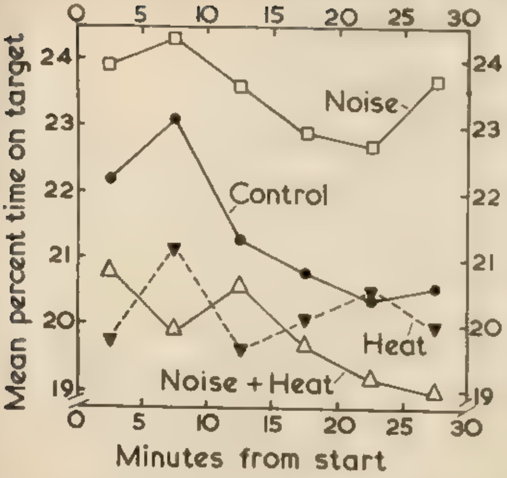


FIGURE 2. The average percent time on target for successive 5-min. periods when tracking with peripheral lights.

Statistical tests. Nonparametric tests (Siegel, 1956) are always used. All tests are two-tailed except where stated.

RESULTS

Heat noise interaction in tracking with peripheral lights. Figure 2 shows that in the control condition the average time on target is less during the last 5 min. of the tracking task than during the first 5 min. ($p < .05$). A similar result is reported in the previous paper (Poulton et al., in press), and by Hockey (1970a, 1970b).

Compared with the control condition, the noise alone produces an improvement which is reliable during the last 5 min. of the task ($p < .05$). The heat alone produces a decrement which is reliable during the first half of the task ($p < .05$). The directions of the changes with noise and heat follow those of previous experiments (Hockey, 1970a, 1970b; Poulton et al., in press).

The novel aspect of the results is the reliable interaction, between the heat and the noise, which is particularly marked during the last 5 min. of the task ($p < .02$). Here heat by itself produces only a small decrement of .5 in the average percent time on target. But when added to the noise, the heat changes an improvement of 3.0 to a decrement of 1.5. The reduced percent time on target produced by adding the heat to the noise is therefor[e] com-pared with only .5 for the heat itself. There are no reliable difference[s] on the lights task related to the noise, [or] to the heat.

Shift in control condition with [the] five-choice task. The five-choice task [was] used also in the previous experiment o[n th]e loss of a night's sleep paired with [the] heat (Poulton et al., in press). It is [perfo]rmed second in both experiments, s[witc]hed between tracking with peripher[al] lights and the vigilance task. The mai[n] differ-ence between the two experiment[s is] that the loss of sleep of the previous ex[per]iment is changed to low-frequency nois[e in] the present experiment. Table 2 sho[ws] that the change has a reliable effect upo[n m]ean percent errors in the control co[ndi]tion. During the last 5 min. of the task, t[he] con-trol mean is 3.2% in the present [ex]peri-ment with noise, compared with 1.[?]% in the previous experiment with loss of [s]leep ($p < .02$).

The bracketed entry at the top [o]f the table shows that the two groups o[f] men are well matched in the main c[o]ntrol practice on the day before the experiments start. Thus the difference between the two control conditions must be due to range effects produced by the balanced experimental design of Table 1 and the comparable design of the previous experi-ment. The particular stresses which are paired together affect the control condi-tion. Since the control condition provides the baseline against which the effects of

TABLE 2

MEAN PERCENT ERRORS ON THE FIVE-CHOICE TASK IN TWO EXPERIMENTS ON PAIRS OF STRESSES

Noise and heat		Loss of sleep and heat[a]	
[Practice control	2.1]	[Practice control	2.3]
Control	2.1[b]	Control	1.4[b,d]
Noise	1.4[c]	Loss of sleep	4.0
Heat	2.9	Heat	2.4
Noise + heat	2.2	Loss of sleep + heat	4.5

[a] Data from Poulton, Edwards, and Colquhoun (in press).
[b] Two control conditions different at .05 level with a one-tailed test.
[c] Noise different from all other conditions at the .05 level.
[d] Control different from all other conditions at the .05 level or better.

the stresses are measured, it means that the measured effect of a stress in an experiment on pairs of stresses depends upon the stress with which it is paired.

The left side of Table 2 shows that in the present experiment the low-frequency noise improves performance, while the mild heat if anything degrades performance. Whereas the right side of the table shows that in the previous experiment both the loss of a night's sleep and the heat degrade performance. Probably the difference between the two range effects is due to the beneficial effect of the low-frequency noise in the present experiment. The noise condition gives 1.4% errors, which corresponds to the 1.4% of the control condition of the previous experiment. Thus in the present experiment the noise condition appears to become the baseline condition, while the control condition becomes a noise-deprived condition.

Other measures of performance obtained with the five-choice task also indicate the beneficial effect of the low-frequency noise. During the first 5 min. there are reliably ($p < .05$) more correct responses in the noise than in the control condition. In addition, combining the noise with the heat reliably ($p < .02$) reduces the number of gaps of 1.5 sec. between responses, compared with the heat alone.

Heat noise interaction in visual vigilance. Figure 3 shows how true and false detections are related during the first and last 7.5-min. periods. For the full 30 min. there are more false detections in the heat than in the control condition ($p < .01$), as in the previous paper (Poulton et al., in press). The proportion of false detections is increased also in the noise, although the difference between the noise and the control condition is not reliable ($p > .05$). When the noise and heat are combined, the proportion of false detections is not far above the control level. The interaction is reliable ($p < .02$). Thus adding the heat to the noise largely eliminates the undesirable increase in false detections which both stresses produce on their own.

The noise gives the greatest proportion of true detections during the first 7.5 min.,

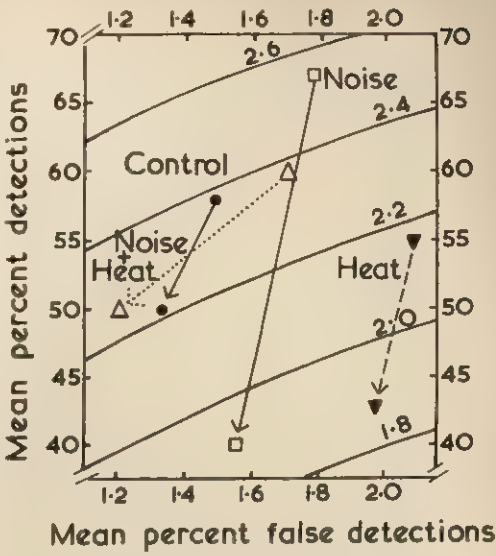


FIGURE 3. The average percent of detections in the visual vigilance task during the first and last 7.5-min. periods, plotted against the average percentage of false detections. (False detections are given as a percentage of the number of nonsignal light flashes. There are 2.2 percent of true signals. The arrows point from the value for the first 7.5 min. to the value for the last 7.5 min. The contours represent equal values of d', an index of the ability to detect signals.)

but changes reliably ($p < .01$) to give the smallest proportion during the last 7.5 min. For the full 30 min., there are no reliable differences in true detections between any of the four conditions. However the noise gives reliably more certain detections than the control condition ($p < .05$, one-tailed test), and reliably fewer moderately certain detections ($p < .01$). The man responds reliably more quickly when he is certain, than when he is uncertain ($p < .02$).

DISCUSSION

Interactions described in terms of arousal theory. Reliable interactions between the heat and the noise are illustrated in Figure 2 for the tracking, and in Figure 3 for false detections in the vigilance. In both cases, the combined effect of the two stresses is smaller than the sum of the two separate effects. In the past such interactions have usually been described in terms of the inverted-U relationship of arousal theory. The effects of the combined stresses will tend to cancel each other if one

stress increases arousal while the other stress reduces arousal. Cancellation can also occur if both stresses increase arousal, and the person thus becomes overaroused (Poulton, 1970, chap. 25).

Noise is generally taken to be an arouser. It produces physiological changes which are associated with arousal. On the five-choice task it reduces the effects of a night without sleep, which presumably lowers the level of arousal (Wilkinson, 1963).

The relationship between heat and arousal is less straightforward. Arousal probably increases while a person is being heated up (Poulton & Kerslake, 1965), and when he is uncomfortably hot (Colquhoun & Goldman, 1972; Wilkinson, Fox, Goldsmith, Hampton, & Lewis, 1964). Whereas a constant mild warmth probably reduces the level of arousal (Wilkinson et al., 1964). In the previous experiment on heat paired with loss of a night's sleep, the heat reduces the detrimental effect of the loss of sleep at the start of the auditory vigilance task (Poulton et al., in press). Thus here the heat must act as an arouser.

The results in Figure 3 for the visual vigilance task can also be described by taking the heat as an arouser. Both the heat and the noise increase arousal and produce more false detections. Adding the heat to the noise produces overarousal, and so cancels the increase in false detections which both stresses produce on their own.

The reliable interaction in Figure 2 for the time on target can also be described in terms of the overarousal which results from combining the two arousers heat and noise. However, the results in Figure 2 cannot all be described in terms of arousal. If the noise increases the time on target by increasing arousal, the heat which also increases arousal should not reduce the time on target as it does. Clearly some other factor in addition to arousal is required to account for the opposite effects of the heat and noise. A similar difficulty is reported in the previous paper where heat is paired with loss of sleep (Poulton et al., in press).

Alternative explanations of interactions. In order to discuss the interactions between stresses in terms of arousal theory, it has to be assumed that an observed interaction reflects a genuine interaction in the underlying process in the brain. However there are other possible causes of an observed interaction. First, the measure of performance may not be related linearly to the underlying brain process. Thus x and z on the left of Figure 1 can be reliably different, without an interaction in the underlying brain process. This possibility is difficult to disprove, because the underlying brain process cannot be observed directly.

The time-on-target measure of Figure 2 is notoriously nonlinear. Bahrick, Fitts, and Briggs (1957) show that on this measure the shape of the learning curve depends upon the size of the target. It would be foolish to claim that performance with the particular size of target used here is more likely to be related linearly to the underlying brain process than is performance with some other size of target. But the criticism of a possible nonlinear relationship to brain function applies to any measure of performance.

A second possible cause of a reliable interaction between two stresses is suggested by the results of Table 2. The relative sizes of x and z on the left of Figure 1 may be distorted by a range effect introduced by a within-subjects experimental design. The present experiment provides no clear evidence on this point, because the range effects appear in the five-choice task, while the reliable interactions appear in the two remaining tasks. But the implications are worth considering.

Many range effects involve a central tendency (Poulton, 1973). Usually, the range is reduced. If the reduction is symmetrical, it need not affect the calculation of the interaction very much. This is because the interaction on the left of Figure 1 is represented by $(x - z)$. If the scale contracts symmetrically about its center, the value of $(x - z)$ may not change much.

But if the range effect has a bias in a single direction, it may, for example, reduce the size of x without appreciably changing the size of z. If so, the difference $(x - z)$ will be smaller than it should be. When the value of $(x - z)$ is negative, as in experiments where the combined effect of two stresses is smaller than the sum of the two separate effects, the range effect will then increase the chances of obtaining a reliable interaction. Both the reliable interactions reported here, and all those reported in the literature, may possibly be produced or exaggerated by an artifact of this kind.

When the value of $(x - z)$ is positive, as in experiments where the combined effect is larger than the sum of the two separate effects, a range effect which reduces the size of $(x - z)$ will reduce the chances of obtaining a reliable interaction. Since no reliable interactions in-

lving potentiation have yet been reported, an artifact of this kind may well be operating in experiments where potentiation is to be expected. Clearly, in order to investigate the interactions between stresses unbiased by range effects, it is necessary to use a separate-groups design. This has not yet been done.

Beneficial effects of low-frequency noise. Table 2 shows that on the five-choice task the low-frequency or green noise reliably reduces the mean percent errors. The noise also reliably increases the rate of responding during the first 5 min. These are entirely new effects of noise on the five-choice task. Presumably the low-frequency noise increases arousal, without having any of the detrimental effects of the noise used in previous experiments.

There are 10 previous experiments where the five-choice task is performed in noise. In all but two of the experiments the noise is white with an approximately flat spectrum. In 7 of the 10 experiments, noise reliably increases errors (Broadbent, 1953, 1957; Colquhoun & Edwards[3]; Hartley, 1973, in press-a, in press-b, Experiment 1; Wilkinson, 1963, Experiment 1). In the remaining three experiments the effect of noise upon errors is not reliable (Corcoran, 1962; Hartley, in press-b, Experiment 2; Wilkinson, 1963, Experiment 2). In none of the 10 experiments is there a reliable reduction in errors in noise, as there is here.

In only one of the 10 experiments does noise have a reliable effect upon the rate of responding. Here the rate of responding is reduced (Colquhoun & Edwards, see Footnote 3). In none of the experiments is the rate of responding reliably increased in noise, as it is at the start of the present task.

The quite different results in the present experiment must be due partly to the spectrum of the noise. There is a 7.5 db. per octave attenuation, which greatly reduces the intensity of the high frequencies. It largely prevents the masking of taps, clicks, and speech, which is produced by the unattenuated high frequencies of white noise. Stevens (1972) suggests that the masking may be responsible for the deterioration in noise, when it occurs.

Broadbent's (1957) experiment partially supports this view. He takes machinery noise having roughly equal energy per octave. He filters it at 2,000 Hz. to give either predominantly high-frequency or low-frequency noise. Both kinds of noise increase errors on the five-choice task, but the low-frequency noise has the reliably smaller effect. Presumably a noise with still greater filtering of the high frequencies, might turn the small increase in errors into the reliable decrease shown in Table 2.

However, the spectrum of the noise cannot alone account for the results in Table 2, because Hartley (1973) reports a reliable increase in errors using noise with a similar spectrum, but at a sound-pressure level of 100 db.(A), or 118 db.(C). Thus both the spectrum and the intensity of the noise are important. To obtain a reduction in errors, the low-frequency noise must not be too loud.

The other beneficial effects of the low-frequency noise are similar to previously reported effects of noise with a flatter spectrum. The improvement in the tracking illustrated in Figure 2 corresponds to the improvement in noise reported by Hockey (1970a, 1970b) toward the end of the tracking task. The effect of the low-frequency noise in increasing confidence in the visual vigilance task, occurs also in visual vigilance with the white noise used by Broadbent and Gregory (1963, 1965).

Most probably, both low-frequency noise at 102 db. (C) and white noise at 100 db. (A) improve performance by increasing the level of arousal. Where the low-frequency noise improves performance while the white noise degrades performance, it may be due to the masking of important auditory cues by the more intense high frequencies of the white noise, as Stevens (1972) suggests.

[3] W. P. Colquhoun & R. S. Edwards. Interaction of noise with alcohol on a task of sustained attention. Study in preparation.

REFERENCES

BAHRICK, H. P., FITTS, P. M., & BRIGGS, G. E. Learning curves—Facts or artifacts? *Psychological Bulletin*, 1957, **54**, 256–268.

BROADBENT, D. E. Noise, paced performance and vigilance tasks. *British Journal of Psychology*, 1953, **44**, 295–303.

BROADBENT, D. E. Effects of noise of high and low frequency on behavior. *Ergonomics*, 1957, **1**, 21–29.

BROADBENT, D. E., & GREGORY, M. Vigilance considered as a statistical decision. *British Journal of Psychology*, 1963, **54**, 309–323.

BROADBENT, D. E., & GREGORY, M. Effects of noise and of signal rate upon vigilance analysed by means of decision theory. *Human Factors*, 1965, **7**, 155–162.

BROWN, I. D. An asymmetrical transfer effect in

research on knowledge of performance. *Journal of Applied Psychology*, 1966, **50**, 118–120.

Colquhoun, W. P., & Goldman, R. F. Vigilance under induced hyperthermia. *Ergonomics*, 1972, **15**, 621–632.

Corcoran, D. W. J. Noise and loss of sleep. *Quarterly Journal of Experimental Psychology*, 1962, **14**, 178–182.

Grether, W. F. Effects on human performance of combined environmental stresses. (Tech. Rep. 70-68) Wright-Patterson Air Force Base, Dayton, Ohio: Aerospace Medical Research Laboratory, 1970.

Hartley, L. R. Effect of prior noise or prior performance on serial reaction. *Journal of Experimental Psychology*, 1973, **101**, 255–261.

Hartley, L. R. Comparison of performance with headphone and free-field noise. *Journal of Experimental Psychology*, in press. (a)

Hartley, L. R. Performance during continuous and intermittent noise and wearing ear protection. *Journal of Experimental Psychology*, in press. (b)

Hockey, G. R. J. Effect of loud noise on attentional selectivity. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 28–36. (a)

Hockey, G. R. J. Signal probability and spatial location as possible bases for increased selectivity in noise. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 37–42. (b)

Leonard, J. A. Five-choice serial reaction apparatus. (Rep. 326) Cambridge, England: Medical Research Council, Applied Psychology Unit, 1959.

Poulton, E. C. *Environment and human efficiency.* Springfield, Ill.: Charles C. Thomas, 1970.

Poulton, E. C. Unwanted range effects from using within-subject experimental designs. *Psychological Bulletin*, 1973, **80**, 113–121.

Poulton, E. C., Edwards, R. S., & Colquhoun, W. P. The interaction of loss of a night's sleep with mild heat: Task variables. *Ergonomics*, in press.

Poulton, E. C., & Kerslake, D McK. Initial stimulating effect of warmth upon perceptual efficiency. *Aerospace Medicine*, 1965, **36**, 29–32.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Stevens, S. S. Stability of human performance under intense noise. *Journal of Sound and Vibration*, 1972, **21**, 35–56.

Wilkinson, R. T. Interaction of lack of sleep with knowledge of results, repeated testing and individual differences. *Journal of Experimental Psychology*, 1961, **62**, 263–271.

Wilkinson, R. T. Interaction of noise with knowledge of results and sleep deprivation. *Journal of Experimental Psychology*, 1963, **66**, 332–337.

Wilkinson, R. T. Sleep deprivation: Performance tests for partial and selective sleep deprivation. In L. A. Abt & B. F. Riess (Eds.), *Progress in clinical psychology.* Vol. 7. New York: Grune, 1969. (a)

Wilkinson, R. T. Some factors influencing the effect of environmental stressors upon performance. *Psychological Bulletin*, 1969, **72**, 260–272. (b)

Wilkinson, R. T., Fox, R. H., Goldsmith, R., Hampton, I. F. G., & Lewis, H. E. Psychological and physiological responses to raised body temperature. *Journal of Applied Physiology*, 1964, **19**, 287–291.

(Received July 11, 1973)

urnal of Experimental Psychology
'74, Vol. 102, No. 4, 629–633

STORAGE CODING TRADE-OFF IN SHORT-TERM STORE[1]

FRANCIS S. BELLEZZA[2] AND RICHARD J. WALKER[3]

Ohio University

Information rehearsed in short-term store (STS) can be primarily new information, if only temporary storage is required; or it can be a mixture of old and new information, if transfer to long-term store is to be optimized. Because STS is of limited capacity, a storage coding trade-off should take place in the use of STS capacity as the memory task changes. Using a single-trial free-recall learning task it was found that *S*s using a storage strategy recalled more words in immediate recall, but fewer words in final free recall, than did *S*s using a coding strategy. Also, *S*s using a storage strategy recalled fewer new words in final free recall than did *S*s using a coding strategy. The results were interpreted as supporting the concept of a storage–coding trade-off in STS.

In dual-storage memory models such as those of Atkinson and Shiffrin (1968) and Waugh and Norman (1965), the process of rehearsal serves to maintain new information in short-term store (STS), and at the same time transfers some of this information with each rehearsal into long-term store (LTS). Recent evidence indicates that rehearsal may not be a unitary process that simultaneously maintains information in STS and transfers it to LTS. Under some conditions, rehearsal may maintain information in STS without transfer to LTS. To support this view, Jacoby and Bartz (1972) have shown that increasing the number of rehearsals in STS does not necessarily lead to more information being transferred to LTS. Although rehearsal in STS may take place without transfer to LTS, the opposite does not seem to be true. Atkinson and Shiffrin (1971) cite evidence which shows that transfer of new information to LTS cannot take place without rehearsal.

The storage and transfer functions of rehearsal are interdependent because both are control processes in STS and both use STS capacity (Atkinson & Shiffrin, 1968). Shiffrin and Atkinson (1969) discuss how rehearsal, under the control of *S*, can be used primarily for maintenance of new information in STS or for transfer of information from STS to LTS. The *S* uses a storage strategy if he fills STS to capacity with new information and rehearses it. He is able to maintain maximum new information in STS, but the rate of transfer to LTS is reduced. The *S* can increase rate of transfer to LTS by using a coding strategy that consists of adding appropriately chosen old information retrieved from LTS to new information in STS and rehearsing the entire ensemble. This description of transfer from STS to LTS implies that the amount of new information being stored in STS must decrease if *S* uses a coding strategy rather than a temporary-storage strategy. Because of the limited capacity of STS (Miller, 1956), *S* must trade off storage space for encoding space, or vice versa, depending on the strategy he wishes to use.

There is some evidence from recall experiments which shows that *S*s may use a strategy involving more encoding into LTS if recall is always delayed rather than immediate. Bartz (1969) presented *S*s with 9-digit lists in a serial-learning task where every third list was identical. All *S*s improved in performance on the identical lists. However, those *S*s in the delayed-recall condition attained a higher performance level than those in the immediate-recall condition. In a multitrial free-recall experi-

[1] This research was supported in part by Grant OURC-437 to the first author from the Ohio University Research Committee.

[2] Requests for reprints should be sent to Francis S. Bellezza, Department of Psychology, Ohio University, Athens, Ohio 45701.

[3] Now at the University of London, London, England.

ment, Bellezza and Dixon[4] alternately presented a list of words and a list of numbers. The *S*s in the immediate-recall condition recalled the items from each list immediately after each list was presented. The *S*s in the delayed-recall condition recalled the other list after each list was presented. After 10 trials, the performance of the delayed-recall group was better on both lists than that of the immediate-recall group. Although it may be that *S*s in the delayed-recall conditions in these 2 experiments were using STS capacity to optimize coding into LTS, there is no way to determine if the amount of new information in STS decreased correspondingly.

The present experiment used a gaming situation in which all *S*s received 1 point for each item immediately recalled. However, *S*s in the coding condition were informed before the words were presented that they would receive 10 additional points for all words that they could recall in a final free recall. The *S*s in the storage condition were not informed of these additional points until shortly before final free recall was requested. It was predicted that *S*s in the coding condition would do more poorly in immediate recall, but would do better in final free recall.

Method

Subjects. The *S*s were 48 volunteers, drawn from a pool of introductory psychology students at Ohio University, who received course credit for their participation.

Procedure. The *S*s were run individually, using a Lafayette IBM memory drum, and were informed that they were in a memory game in which they were to try to win as many points as they could. All *S*s were told that they would be presented 7 lists of numbers and would receive 1 point for each number that they could free recall immediately after each list was presented, as long as they did not pause for more than 2 sec. between responses. A pause of more than 2 sec. was used to indicate that *S* had emptied STS (Murdock & Okada, 1970). Each *S* was then presented 7 lists of 10 2-digit numbers, with both digits different and no zeros used. All 10 numbers on each list were presented at once in a row for 15 sec. The *S*s recalled orally, and *E* checked items off a score sheet as they were recalled. As soon as *S* stopped recalling, the next list was presented. After the lists of numbers were presented, the 24 *S*s in the coding condition were informed that they would be presented 7 lists of words and would again receive 1 point for each word immediately recalled. In addition, these *S*s were told that there would be a final free recall in which they would receive 10 additional points for each correctly recalled word from the 7 previous lists. The other 24 *S*s in the storage condition were not informed of the final free recall. Each *S* was then presented 7 lists of 10 1-syllable words randomly drawn from Thorndike and Lorge (1944) and having a frequency of occurrence of 10–40 per million. The immediate-recall procedure for words was the same as the procedure used for numbers. Each number and word was used only once with each *S*. Eight different randomizations of the lists of numbers and words were devised, and 6 *S*s were run on each form. After all 7 lists of words were presented, all *S*s were presented with a 3-digit number and were asked to count backward by 7s as fast as they could for 1 min. This was to eliminate any items in STS. The *S*s were then requested to try to recall as many of the words in the 7 lists as possible, and *S*s in the storage condition were then informed of the 10 additional points for each word correctly recalled. Recall was oral, and *S*s were given 1 min. and 50 sec. for final free recall.

Analysis. The results of the experiment were evaluated using a $2 \times 8 \times 7$ analysis of variance, with the specific factors being strategy condition, randomization form, and list position. Strategy condition and randomization form were between-*S*s factors. The analysis was done on 3 different dependent variables: the number of 2-digit numbers per list immediately recalled, the number of words per list immediately recalled, and the number of words per list recalled in final free recall.

Results

The analysis of variance of the number of 2-digit numbers recalled showed only the significant main effect of list position, F (6, 192) = 4.02, $p < .001$, $MS_e = 1.15$. Figure 1 shows the number of numbers immediately recalled from each list. There seems to be a slight interlist primacy effect, and a trend analysis (Kirk, 1968) showed the linear component to be significant, F (1, 192) = 6.24, $p < .025$, $MS_e = 1.15$. Because there were no significant effects involving strategy condition, it can be concluded that *S*s were performing at the same level in immediate recall before different STS strategies were induced.

The analysis of variance on the number of words immediately recalled per list

[4] F. S. Bellezza and R. W. Dixon. Storage versus coding functions of short-term store. Paper presented at the meeting of the Midwestern Psychological Association, Cleveland, May 1972.

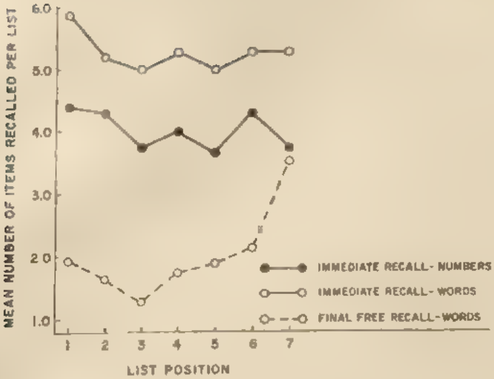


FIGURE 1. Number of items recalled from each 10-item list.

showed a significant main effect of strategy condition, F (1, 32) = 62.51, $p < .001$, $MS_e = 2.56$. The *S*s in the storage condition recalled a mean of 6.00 words/list and *S*s in the coding condition recalled 4.64 words/list. These results are shown in Figure 2. There was also a significant effect of list position, F (6, 192) = 3.92, $p < .001$, $MS_e = 1.17$, as shown in Figure 1. As in the immediate recall of numbers, there appears to be an interlist primacy effect. A trend analysis showed the linear component to be marginally significant, F (1, 192) = 3.67, $p < .10$, $MS_e = 1.17$. The one other significant source of variation was the Strategy Condition × Randomization Form interaction, F (7, 32) = 2.39, $p < .05$, $MS_e = 2.56$. Inspection of the strategy condition means in each randomization form showed that *S*s in the storage condition performed better than did *S*s in the coding condition on all forms except the fourth, where the means were the same. There is no obvious explanation for this except individual differences among *S*s.

When the data from the final free recall were analyzed, the main effect of strategy condition was again significant, F (1, 32) = 46.01, $p < .001$, $MS_e = 1.91$. As shown in Figure 2, *S*s in the storage condition recalled 1.48 words/list, whereas *S*s in the coding condition recalled 2.48 words/list. There was a significant main effect due to list position, F (6, 192) = 13.08, $p < .001$, $MS_e = 1.89$. As shown in Figure 1, not only is there a slight interlist primacy effect, but there is also a pronounced interlist recency effect, even though *S*s engaged in a difficult 60-sec. interpolated task before final recall began. No other sources of variation were significant in this analysis.

To determine to what degree the same words were given in immediate and in final free recall, all the words on the lists were placed into 1 of 4 categories: those given in both immediate and final free recall; those given in immediate recall but not in final free recall; those not given in immediate recall but given in final free recall; and those not recalled at all. The mean numbers of words in each of these 4 categories recalled by each S in each strategy condition are given in Table 1. In order to determine if the strategy conditions differed in all 4 categories of recall, a 2 × 8 × 4 analysis of variance was performed, with the specific factors being strategy condition, randomization form, and recall category. Strategy condition and randomization form were between-*S*s factors. The dependent variable was the total number of words recalled by each S in each category. The Strategy Condition × Recall Category interaction was significant, F (3, 32) = 28.21, $p < .001$, $MS_e = 29.03$. Tests of simple main effects (Kirk, 1968) showed that the

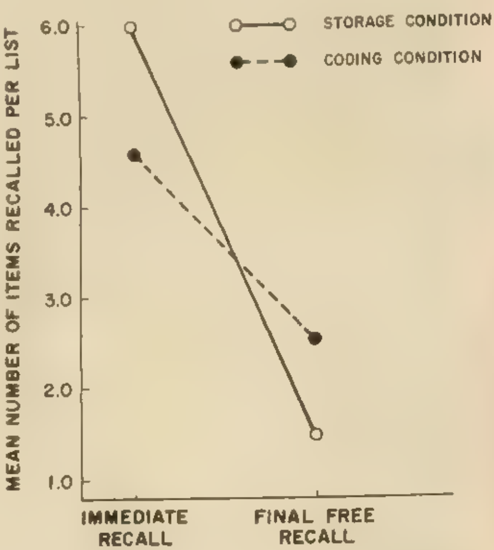


FIGURE 2. Number of words recalled per list in the storage and coding conditions.

TABLE 1

Mean Number of Words Recalled in Each Category in Final Free Recall

Condition	Category			
	IF	I$\overline{F}$	$\overline{I}$F	$\overline{IF}$
Coding	11.75	20.71	5.63	31.91
Storage	9.00	33.00	1.38	26.62
Difference	1.75*	−12.29**	4.25**	5.29**

Note. Abbreviations: IF = words given in both immediate and final free recall; I$\overline{F}$ = words given in immediate but not in final free recall; $\overline{I}$F = words given in final but not in immediate free recall; and $\overline{IF}$ = words not recalled at all.

* $p < .05$.
** $p < .001$.

means of the strategy conditions were significantly different in each recall category.

Discussion

The results support the concept of a storage–coding trade-off in STS. The trade-off is necessary because STS is of limited capacity. If *S* desires to transfer new information into LTS, he must load stored information into STS so that the new information can be appropriately placed in LTS (Shiffrin & Atkinson, 1969). This reduces the amount of new information that can be stored in STS. Because of the gaming situation used in the present experiment, *S*s in the storage condition loaded STS with as much new information as possible and stopped rehearsing it only after immediate recall. As shown in Figure 2, the storage group recalled significantly more words in immediate recall, but significantly fewer words in final free recall, than did the coding group. This indicates that STS capacity was being used in different ways by each group. It can be seen from Table 1 that the storage group recalled 1.38 words in final free recall that were not recalled in immediate recall. In the coding group, this number was 5.63. With a coding strategy, *S* would be likely to stop rehearsing an item when he thought it was coded into LTS, so that other words could be coded. This means that these coded words would not be available in the immediate-recall condition, which required almost uninterrupted responding. With the greater amount of time given for search and retrieval in final free recall, these words could be recalled. The *S*s in the storage condition were likely not to stop rehearsing any word transferred to LTS because this word would not be available for immediate recall. The results shown in [illegible] predicted from the notion tha[illegible] strategy necessarily consists of ad[illegible] appropriately chosen old information retri[illegible]d from LTS to new information in STS and re[illegible]ersing the entire ensemble. Other interpreta[illegible]ns of the data are possible. For example, a[illegible]ing strategy that involved nothing more th[illegible]lectively rehearsing a fewer number of ite[illegible]ould lead to the same results.

The better immediate recall perform[illegible] in the early lists of numbers and of word[illegible] in Figure 1 is probably not the result of [illegible] Bartz (1969) found an overall increase [illegible] the number of digits recalled as a function [illegible] position, although there was an initial de[illegible] in his first 3 lists. Melton (1963) also fou[illegible] a practice effect in a serial recall task. An al[illegible]nate explanation is that first-list performa[illegible] in immediate recall has a larger LTS co[illegible]ponent than does performance on later list[illegible] On the first list, *S* may be able to retrieve in[illegible]formation quickly from LTS, since information from only one list is stored there, and the problems of search and decision are easier and therefore faster (Shiffrin, 1970). The interlist primacy effect which seems to have occurred in final free recall could be the result of the earliest lists having more temporally distinct associated cues and hence being more accessible in LTS (Shiffrin, 1970). However, the effect could also be the result of the better immediate recall of the early lists, since the recall of an item has the same effect as an additional presentation trial (Tulving, 1967). The mechanism of memory search developed by Shiffrin can account for the interlist recency effect found in the final free recall, because information from the more recent lists is more accessible in LTS.

The results of the present experiment indicate that the concept of a storage–coding trade-off in STS could be a useful one in short-term memory experiments

References

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Vol. 2. Advances in research and theory.* New York: Academic Press, 1968.

Atkinson, R. C., & Shiffrin, R. M. The control of short-term memory. *Scientific American*, 1971, 225, 82–90.

Bartz, W. H. Repetition and the memory stores. *Journal of Experimental Psychology*, 1969, 80, 33–38.

Jacoby, L. L., & Bartz, W. H. Rehearsal and transfer to LTS. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 561–565.

Kirk, R. E. *Experimental design procedures for the behavioral sciences.* Belmont, Calif.: Wadsworth, 1968.

Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 1–21.

Miller, G. A. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

Murdock, B. B. & Okada, R. Interresponse times in single-trial free recall. *Journal of Experimental Psychology*, 1970, **86**, 263–267.

Shiffrin, R. M. Memory search. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

Shiffrin, R. M., & Atkinson, R. C. Storage and retrieval processes in long-term memory. *Psychological Review*, 1969, **76**, 179–193.

Thorndike, E. L., & Lorge, I. *The teacher's word-book at 30,000 words.* New York: Colombia University, Teachers College, Bureau of Publications, 1944.

Tulving, E. The effects of presentation and recall of materials in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 175–184.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

(Received July 11, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 634-639

EFFECT OF TEST STIMULUS RANGE ON STIMULUS GENERALIZATION IN HUMAN SUBJECTS[1]

GARY HANSEN, ARTHUR TOMIE, AND DAVID R. THOMAS[2]

University of Colorado

DORIS H. THOMAS

Colorado Women's College

In Experiment I, *S*s were exposed to a 55° line angle training stimulus and then tested for generalization (recognition) with lines of 40°, 55°, and 70° (Group 1); 40°, 47.5°, 55°, 62.5°, and 70° (Group 2); or 25°, 40°, 55°, 70°, and 85° (Group 3). Groups 1 and 2 yielded similar decrements to 40° and 70° lines, indicating no units effects; however, Group 3 responded more to these 2 stimuli, indicating a reliable range effect.

In Experiment II, the training stimulus was a 530-nm. light, and the test values were 515, 530, and 545 nm. (Group 1); 515, 522.5, 530, 537.5, and 545 nm. (Group 2); or 500, 515, 530, 545, and 560 nm. (Group 3). Group 3 responded more to 515 and 545 nm. than did Groups 1 and 2, which did not differ, again indicating a range effect and no units effect. Prior evidence of a units effect was probably due to a confounding of number of units with range of test stimuli.

The amount of stimulus generalization decrement in human *S*s has been shown to be influenced by a number of factors in addition to the physical distance between training and test stimuli. Mednick and Freedman (1960) have hypothesized that one such factor is the number of stimulus units that separate the test stimulus from the training stimulus. Their "units hypothesis" generates 2 primary predictions. (*a*) The amount of generalization decrement between a training stimulus and a given test stimulus along the same dimension is a positive function of the number of intervening test stimuli. (*b*) The generalization decrement obtained for a given number of units away from the training stimulus will be the same for units of different size. Note that Prediction *b* assumes that the number of units is the sole determinant of generalization decrement, whereas Prediction *a* assumes that it is a determinant, but not necessarily the onl

Thomas and Hiss (1963) reported a study designed to test these predictions using wavelength generalization. Three groups of human *S*s viewed a training stimulus of 530 nm. followed by a 5-stimulus generalization test in which the training stimulus occupied the physical midpoint of the test range. The groups varied in the sizes of the test unit, i.e., the physical distance between adjacent stimuli, which were 5, 10, or 20 nm., respectively.

The results reported by Thomas and Hiss (1963) supported the first of the Mednick and Freedman (1960) predictions. Responses to common test values (i.e., those experienced by 2 different groups) decreased as the number of test stimuli between them and the training stimulus increased. This study, however, failed to support the second prediction. A comparison of generalization gradients plotted in terms of number of units revealed a significant effect of unit size. Thomas and Hiss concluded, therefore, that both the number of units and the physical distance between stimuli influenced generalization slope.

The apparent finding of a units effect in the Thomas and Hiss (1963) study is

[1] This research was supported by National Institutes of Health Research Grant HD-03486 and Training Grant MH-10427 under the direction of the third author. Experiment I was performed by the first and second authors. Experiment II was performed by the third and fourth authors while on leave at the University of Iowa. The authors wish to acknowledge the assistance of James L. Thomas in the analysis of the data of Experiment II.

[2] Requests for reprints should be sent to David R. Thomas, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

subject to an alternative interpretation, however. Because the number of test stimuli was held constant (at 5) across groups, unit size was confounded with the range covered by the test series. An increase in unit size produced a proportional increase in the series range. This increase in range could have had an effect on generalization decrement independent of the change in unit size. It is also possible that the units effect is dependent on an interaction between unit size and test range.

A recent experiment by Tomie (1972) suggests that the range of test stimuli used by Thomas and Hiss (1963) may have indeed been a significant factor. Two groups in the Tomie study were exposed to a training stimulus of 505 nm. and were then tested for generalization to values of both longer and shorter wavelengths than the standard. For one group, only a single test value of shorter wavelength (500 nm.) was used; for the other group, the test stimuli included 500, 495, 490, and 485 nm. Thus, for both groups, 500 nm. was one unit removed from the training value but, for one group, there were additional stimuli still farther removed. This group showed significantly more responding to the 500-nm. stimulus, indicating that increasing the range of generalization test stimuli flattens the gradient.

Two investigations of the units hypothesis have been conducted using pigeons and wavelength generalization tests. Friedman (1963) employed a selection of test values similar to that used by Thomas and Hiss (1963), in that he varied unit size while holding the number of test stimuli constant. This produced a unit size by stimulus range confounding, as noted above. Despite this confounding, Friedman found no support for either prediction based on Mednick and Freedman's (1960) units hypothesis. While no reliable differences were found in responding to common test values, the group tested over the widest range *did* respond at the highest level in all 4 relevant comparisons.

An experiment by Marsh (1967) is somewhat difficult to interpret because of variations in level of responding to the training stimulus among the different groups. Marsh used varying numbers of test stimuli, all placed to one side of the training value. He reported, in agreement with Thomas and Hiss (1963), that generalization decrement varied with both the physical size and the number of units separating the training and test stimuli. Although Marsh's experiment was designed in a way which permitted the investigation of a range effect, he found none.

The purpose of the present study was to separate the potential influences of range and size of stimulus unit on generalization decrement in an experiment with human *S*s. Three test conditions were employed involving (*a*) small units–small range, (*b*) large units–large range, and (*c*) large units–small range. The latter condition was achieved by the use of a reduced number of different test stimuli. This produced a group difference in the total number of stimulus presentations, but subsequent analysis indicated that this difference did not affect the slope of the obtained gradient. In order to enhance the generality of the present investigation, both line angle and wavelength test dimensions were employed in 2 separate experiments.

Experiment I

Method

Subjects. The *S*s were 30 male and 30 female undergraduate students enrolled in introductory psychology courses at the University of Colorado.

Apparatus. The apparatus contained a 4.5-cm.-diam. circular aperture in a 122 × 76 cm. black panel. The aperture was covered with translucent plastic. Line angles were produced by backlighting (with a 7.5-w. lamp) a .3-cm.-wide black plastic strip mounted immediately behind and bisecting the aperture cover. Thus, the stimuli were black lines at various angles viewed against a homogeneous white field. The line angles used in this experiment varied from 25° to 85° (counterclockwise from horizontal) in 7.5° steps. Angle adjustment and stimulus backlighting were performed manually by *E*, who was seated behind the panel. The *S*s were seated with their faces approximately 60 cm. from the panel, with the circular aperture approximately at eye level.

Procedure. The *S*s were randomly assigned to 3 groups with the restriction that each group contain 10 males and 10 females. Instructions were the same for all *S*s and included a general explanation of the experimental task. The *S*s were instructed

to study the sample stimulus carefully and to respond on each trial by judging the test stimulus as *same* or *different* in relation to the sample. The instructions emphasized that *S* should respond on each trial and that responses should occur during the stimulus presentation.

All *Ss* received the same sample stimulus (a black line 55° counterclockwise from horizontal) for 15 sec. The generalization test consisted of 9 blocks of trials with the appropriate test series randomized within each block. During the generalization test, each stimulus presentation was 5 sec. in duration, with an intertrial interval of approximately 5 sec. The *E* supplied no feedback to *S* during the generalization test.

Group 1 received stimulus lines of 40°, 55°, and 70° during each test block. Group 2 received 40°, 47.5°, 55°, 62.5°, and 70° lines; and Group 3 received 25°, 40°, 55°, 70°, and 85° lines. The selection of these stimulus values allows an evaluation of the units effect with range held constant (Groups 1 and 2) and an evaluation of the effect of range with the unit size held constant (Groups 1 and 3).

Results and Discussion

Initially a Group × Stimuli (40°, 55°, and 70°) × Sex analysis of variance was performed. Because the variable of sex of *S* neither yielded a significant main effect nor entered into any significant interaction all subsequent comparisons were pool across sex.

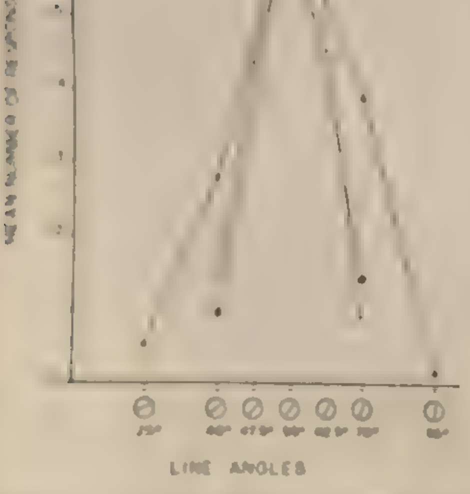


FIGURE 1. Mean generalization gradients for the 3 groups of *Ss* in Experiment I.

The mean numbers of *same* responses f the 3 groups are shown in Figure 1. The data of primary interest are the number of *same* responses to the common stimuli 40°, 55°, and 70°. All gradients peaked essentially the same level. Mean resp to 55° were 6.3, 6.6, and 6.8, for Group 1, 2, and 3, respectively, ($F < 1$).

Same responses to 40° and 70° pr the primary measure for comparing effects of range and unit size on stimu generalization. The gradients of Gro 1 and 2 provide a direct test of the fi prediction generated by the units hypo esis. If generalization decrement var positively with the number of units inte vening between training and test value Group 2 should show fewer *same* responses to 40° and 70° than Group 1. The mean total *same* responses to 40° and 7 these 2 groups were 2.40 and 2.4 tively ($F < 1$). Inspection of Fig shows the gradients produced by th 2 groups to be nearly identical.

The second prediction generated by the units hypothesis is that the generalization decrement shown for a given number of units away from the training stimulus will be the same for units of different size. If this prediction were correct, the level of responding to 40° and 70° for Group 1 should equal that to 47.5° and 62.5° for Group 2. This is clearly not the case. Mean total responses to these respective sets of stimuli were 2.45 and 8.40, which differ significantly, $F(1, 38) = 56.11$, $p < .01$. Another comparison relevant to this prediction involves Stimuli 25° and 85° for Group 3 vs. 40° and 70° for Group 2. Here the mean total responses, .25 and 1.15, respectively, again differ significantly, $F(1, 38) = 5.11$, $p < .05$.

Group 3 shows a substantial level of responding to 40° and 70°, with a mean of [illegible]

$= 10.18$, $p < .01$. Separate Newman-Keuls comparisons between Group 3 and Groups 1 and 2 were also significant ($p < .01$). As is shown in Figure 1, the effect of series range is not dependent on [illegible] to the [illegible] end values. The mean total responses to 25° and 85° were only .25.

It should be noted that the generalization test for Group 1 contained 27 stimulus presentations, compared to 45 for Groups 2 and 3. This procedural difference can be shown to be inconsequential in 2 ways. First the gradients of Groups 1 and 2 were comparable, despite the difference in length of the test. Second, the data for Groups 2 and 3 were reanalyzed so as to include only the first 5 test series (25 stimulus presentations). The gradients were then computed, based on the percentage of opportunities on which a response actually occurred. All of the group comparisons reported for data for the entire test were also observed in this reanalysis, at the same levels of confidence.

This experiment, then, offers no support for the units hypothesis and suggests that the data presented by Thomas and Hiss (1963) also failed to support it. It should be noted that our Groups 2 and 3 varied in the same [illegible] in their study. The gradients show the same direction of differences in responses to the common test values. The inclusion of our Group 1, however, suggests that such differences are the result of test series [illegible] in testing.

EXPERIMENT II

Because the results of Experiment I and of the Tomie (1972) study suggest the [illegible] able, and [illegible] procedure [illegible] work revealed that when the monochromatic lights were matched for brightness (which Thomas and Hiss did not do), a [illegible] was required than the 5 nm. they used. Another change was the use of verbal *same-different* responses instead of the finger-lift response used by Thomas and Hiss. Pilot work in our laboratory has shown the verbal and motor responses to be functionally equivalent. We have also determined that a much shorter exposure to the training stimulus then the 60 sec. used by Thomas and Hiss produces comparable data.

Method

Subjects. The *Ss* were 30 male and 30 female undergraduate students enrolled in introductory [illegible] courses at the University of Iowa.

Apparatus. The apparatus contained a 2.22-cm diam circular aperture in a 76.20-cm. wide by 61.40 cm high black panel. The aperture was covered with translucent plastic. The light source, a Bausch and Lomb Model 33 86 02 monochromator equipped with a tungsten light source, was mounted 22.86 cm behind and at the same height as the circular aperture. A pair of circular Kodak Type M carbon neutral density optical wedges was mounted on the monochromator between the light source and the diffraction grating. All stimuli were equated for brightness by adjusting the settings of the variable [illegible] prior to the initiation of the experiment. Adjustment of both the monochromator and the wedge was done manually by *E* during the intertrial intervals. The *Ss* were seated with their faces approximately 60 cm. from the panel, with the circular aperture approximately at eye level.

Procedure. The *Ss* were randomly assigned to 3 groups with the restriction that each group con[illegible] of the [illegible] to study the sample stimulus carefully and to [illegible] test stimulus in relation to the sample. The [illegible] trial and that responses should occur during the stimulus presentation.

All *Ss* received the same sample stimulus, a 530-nm. light, for 15 sec. The generalization test consisted of 9 blocks of trials, with the appropriate test series randomized within each block. During the generalization test, each stimulus presentation was 5 sec. in duration, with an intertrial interval of [illegible] to *S* during the generalization test.

Stimuli received were of 515, 530, and 545 nm. for Group 1, 515, 522.5, 530, 537.5, and 545 nm

for Group 2; and 500, 515, 530, 545, and 560 nm. for Group 3. The selection of these stimulus values allowed an evaluation of the units effect with range held constant (Groups 1 and 2) and an evaluation of the effect of range with unit size held constant (Groups 1 and 3).

Results and Discussion

Because the rationale and design of Experiment 2 parallel Experiment I, the presentation of results will also be parallel, and can therefore be somewhat condensed. Again, an initial Group × Stimuli (515, 530, and 545 nm.) × Sex analysis of variance was performed. Because the variable of sex of *S* neither yielded a significant main effect nor entered into any significant interaction, all subsequent comparisons were pooled across sex. The generalization gradients of the 3 groups are presented in Figure 2. The 3 groups responsed quite comparably to the 530-nm. training stimulus value, with 8.1, 7.1, and 7.5 responses for Groups 1, 2, and 3, respectively ($F < 1$). The units hypothesis generates the prediction that Group 2 should show fewer *same* responses to 515 and 545 nm. than Group 1. The mean total *same* responses to these 2 values were 2.75 and 3.45 for Groups 1 and 2, respectively. This difference is insignificant ($F < 1$) and in the opposite direction from that predicted. A second prediction generated by the units hypothesis is that Group 1 should respond as much to 515 and 545 nm. as Group 2 does to 522.5 and 537.5 nm. Mean responses to these respective sets of stimuli were 2.75 and 7.40, which differ significantly, $F(1, 38) = 32.79$, $p <$ .01. According to the same principle, Group 3 should respond as much to 500 and 560 nm. as Group 2 does for 515 and 545 nm. However, the mean total responses, .25 and 3.45, respectively, again differ, $F(1, 38) = 21.53$, $p < .01$.

FIGURE 2. Mean generalization gradients for the 3 groups of *S*s in Experiment II.

Group 3 showed a substantial level of responding to 515 and 545 nm., with a mean of 5.70 total responses to these 2 stimuli. Analysis of variance including all 3 groups showed a significant difference in summed responses to 515 and 545 nm., $F(2, 57) = 4.85$, $p < .025$. Separate Newman–Keuls comparisons between Group 3 and Groups 1 and 2 were also significant ($p < .05$). As is shown in Figure 2, the effect of range of test stimuli is not dependent on any substantial responding to the added end values. Indeed, responses to these values almost never occur.

Finally, a reanalysis of the data including only the first 5 test series of Groups 2 and 3 yielded all of the same group differences and similarities reported above for gradients obtained over the entire course of testing. Thus, in all major respects, the results obtained with wavelength stimuli in Experiment II parallel those obtained in Experiment I with line angles.

General Discussion

Several interpretations may be offered for the effect of the range of test stimulus values on the slope of stimulus generalization gradients. Capehart, Tempone, and Hebert (1969) have proposed that Adaptation level (AL), rather than the absolute value of the training stimulus, serves as the reference value in stimulus generalization studies. This view assumes that the training stimulus becomes the AL value and, in subsequent generalization

testing manipulations which affect the AL value affect generalization accordingly. Such manipulations include the use of a test series asymmetrically spaced around the training value, the over representation of certain test values during testing, etc. In the present experiment, the range of test stimuli in a symmetrical distribution around the training stimulus would presumably not affect the mean AL, but it would affect AL variance and thus *S*'s degree of certainty about whether a particular test stimulus was at the AL value at that time. If *S* had a set to "respond when in doubt," then increasing the variance of AL could result in a flattened gradient.

An alternative interpretation involves an application of Parducci's (1965) range-frequency hypothesis, originally developed to predict performance in magnitude estimation tasks. If *S*s placed the line angle stimuli (for example) in 3 categories (sharper than the training stimulus, same as the training stimulus, or flatter than the training stimulus), then broadening the range of test stimuli might extend the boundaries of the *same* category, with a consequent flattening of the line angle gradient. The present study does not permit a rejection of either of these interpretations, and there may be others which are equally tenable.

With the proposed reinterpretation of the Thomas and Hiss (1963) study, it appears that support for Mednick and Freedman's (1960) units hypothesis is meager at best. None of the studies explicitly designed to test the hypothesis have provided confirmatory evidence. Furthermore, the evidence cited in the original Mednick and Freedman article was entirely based on a comparison of experiments performed in different laboratories, such that innummerable procedural differences might have accounted for the results obtained.

Evidence for a range effect is somewhat more substantial, having been found reliably in a wavelength generalization study by Tomie (1972) and in the present line angle and wavelength experiments. In neither of the wavelength generalization studies with pigeons by Marsh (1967) and Friedman (1963) was reliable evidence found for a range effect, although in the latter case, there was some supportive evidence, which failed to achieve statistical significance. The variables that determine the presence or magnitude of a range effect in stimulus generalization have yet to be determined. Capehart et al. (1969) considered and rejected the possibility that infrahuman *S*s may not be subject to contextual effects in generalization testing. We agree with their judgment that, in principle, such effects should not be limited to humans. Giurintano (1972) has demonstrated in a study with pigeons that when training has been very brief, generalization testing with a series of stimuli with all values to one side of the training stimulus may result in a shift of maximal responding to a more centrally located stimulus. Perhaps the amount of training administered to *S*s will similarly prove to be a crucial variable in the investigation of the range effect in stimulus generalization, and what now appears to be a species difference will turn out to be only a procedural one.

REFERENCES

Capehart, J., Tempone, V. J., & Hebert, J. A. A theory of stimulus equivalence. *Psychological Review*, 1969, **76**, 405–418.

Friedman, H. Wavelength generalization as a function of spacing of test stimuli. *Journal of Experimental Psychology*, 1963, **65**, 334–338.

Giurintano, L. P. Stimulus generalization in humans and pigeons as a function of training and test stimulus parameters. Unpublished dissertation, University of Colorado, 1973.

Marsh, G. D. Inverse relationship between discriminability and stimulus generalization as a function of number of test stimuli. *Journal of Comparative and Physiological Psychology*, 1967, **64**, 284–289.

Mednick, S. A., & Freedman, J. L. Stimulus generalization. *Psychological Bulletin*, 1960, **57**, 167–200.

Parducci, A. Category judgment: A range-frequency model. *Psychological Review*, 1965, **72**, 407–418.

Thomas, D. R., & Hiss, R. H. A test of the units hypothesis employing wavelength generalization in human subjects. *Journal of Experimental Psychology*, 1963, **65**, 59–62.

Tomie, A. Adaptation level as a factor in human wavelength generalization. Unpublished master's thesis, University of Colorado, 1972.

(Received July 11, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 640–647

SOURNESS OF ACID MIXTURES [1]

HOWARD R. MOSKOWITZ [2]

Pioneering Research Laboratory, U.S. Army Natick Laboratories, Natick, Massachusetts

The sourness of 5 acids (citric, hydrochloric, phytic, succinic, and gluconolactone) was assessed at 4 concentrations of each acid and in 16 mixtures with citric acid by the method of magnitude estimation. Power functions fit to the sourness of unmixed acids were either added directly to predict mixture sourness, or both acids in the mixture were first converted to equivalent levels of citric acid and then exponentiated according to the appropriate power function. Summation of sourness functions accounted best for mixtures of citric acid with gluconolactone. Summation of acid equivalents was better for mixtures with phytic acid. Both models predicted mixtures with succinic and hydrochloric acids equally well. Mixtures with hydrochloric acid were synergistic. The results are discussed in terms of possible mechanisms underlying taste mixtures.

For the past 50 years, sensory additivity of taste and smell has concerned investigators, who have attempted to derive relations between the perceived intensity of mixture components tasted or smelled in in the unmixed (pure) state and the intensity of the mixture. Three major classes of models have been proposed, two from studies of taste and one from studies of smell. The present article concerns these models as they apply to the sourness of acid mixtures.

Model 1, the summation of concentrations, derives from the earliest studies of taste mixtures and can be best seen in historical perspective in the extensive sweetness studies of Cameron (1947). Cameron noted that the sweetness of mixtures containing 2 or more sugars could be predicted if each sugar component were first compared to a reference sweet substance. In the unmixed state, each component was compared to glucose and the "glucose equivalent" determined experimentally. The sum of the glucose equivalents of the 2 (sometimes 3) components approximated the glucose equivalent of t mixture.

Cameron's (1947) additivity model wa limited to equal-sweetness matches, since, at that time, no appropriate measures were developed for the direct assessment of sensory intensity that would, simultaneously permit subjective magnitudes to be added, subtracted, or multiplied by a constant With the development of magnitude estimation, *S*s have been given a tool to repor sensory intensities along a ratio scale. In this procedure, *S*s assign numbers to taste intensity (S), and these numbers are regressed against concentration (C) (Moskowitz, 1970). The results for acids, evaluated in the unmixed state, suggest that the simple power function $S = kC^n$ provides a reasonable description of the growth of sourness with concentration (Moskowitz, 1971). For 2 acids in a mixture, 2 separate power functions may be obtained:

$$\begin{aligned} \text{Acid A:}\quad S_a &= k_a C_a{}^m \\ \text{Acid B:}\quad S_b &= k_b C_b{}^n. \end{aligned} \qquad [1]$$

The exponents m and n need not be equal, and for organic acids they typically lie between .5 and 1.0, so that sourness is a decelerating function of concentration. The intercepts k_a and k_b may also vary, and reflect the sourness estimate at 1.0 gm. molecular weight (M) of each acid.

Cameron's (1947) additivity model may be expanded to account for power functions

[1] The author wishes to thank Harry L. Jacobs for numerous helpful discussions about the models of additivity and for suggestions about relating psychophysical measurement to electrophysiological studies.

[2] Requests for reprints should be sent to Howard R. Moskowitz, Pioneering Research Laboratory, U.S. Army Natick Laboratories, Natick, Massachusetts 01760.

of sensory intensity. Equation 1 can be solved for equivalents of 1 acid (e.g., Acid A)

$$S_a = S_b \text{ implies } k_aC_a^m = k_bC_b^n, \quad [2]$$

$$C_a = (k_bC_b^n/k_a)^{-m}. \quad [3]$$

Equation 3 presents the concentration of Acid A that is as sour as the corresponding concentration of Acid B. In a mixture of the 2 acids ($C_a + C_b$) the sourness can be predicted by Equation 4:

$$S_{ab} = k_a[C_a + (k_bC_b^n/k_a)^{-m}]^m. \quad [4]$$

Model 2, the summation of perceived sournesses, suggests that the gustatory system treats the 2 acid concentrations as separate inputs, processes them separately, and adds the outputs ("sensory judgments") together. Equation 5 presents Model 2:

$$S_{ab} = k_aC_a^m + k_bC_b^n. \quad [5]$$

Model 2 is simpler than Model 1, since it does not require any cross terms in the mixture equation. It also suggests a substantially different process by which the gustatory system adds together information from a compound source.

The third model, the vector addition of sourness judgments, is derived from an approach to odor mixtures reported by Berglund, Berglund, Lindvall, and Svensson (1972). They proposed that binary mixtures of odors, e.g., dimethyl disulfide + pyridine, hydrogen sulfide + pyridine, etc., may add together together in intensity as if they behaved like vectors separated by a constant angle Q. Their additivity equation is shown in Equation 6:

$$S_{ab} = [S_a^2 + S_b^2 + 2S_aS_b(\cos Q)]^{.5}. \quad [6]$$

When the 2 odors (or here, acids) are qualitatively identical, then the angle separating them, Q, becomes 0, and its cosine becomes 1.0. Note that the vector additivity is expressed in terms of subjectively perceived intensity. A slight modification is needed to represent Equation 6 as the sum of power functions.

A previous study with sweeteners and their mixtures failed to distinguish between Models 1 and 2, viz., taste summation of concentrations vs. summation of subjective intensities (Moskowitz, 1973). In that study, glucose was mixed with fructose, saccharin, or cyclamate, and *S* was asked to estimate perceived sweetness. The 2 models were shown to differ maximally when 2 conditions prevailed: (*a*) the exponents *m* and *n* differed widely from 1.0, and (*b*) as little "noise" as possible was present in the mixture estimates. The 2 models become increasingly similar as the exponents approach 1.0 (at which point they become identical).

The present study concerns acid mixtures and the 3 models of summation. Previous studies with the sourness of organic acids (Beebe-Center, 1949; Moskowitz, 1971) suggested that the rated sournesses of tartaric acid, citric acid, and a number of other compounds grow according to a simple power function of concentration: $S = kC^n$. The exponent *n* is usually lower than 1.0, lying around .7 for citric acid and perhaps slightly higher for tartaric acid. When the measure of concentration is defined in terms of pH ($-\log$ H+ ions in solution, a measure of hydrogen ion concentration), the exponent *n* becomes -1.7, suggesting that apparent sourness grows as an accelerating function.

Method

In 4 experiments, citric acid was evaluated both in unmixed solutions and when mixed with another acid (either hydrochloric, acetic, phytic or gluconolactone). Four concentrations of citric acid and 4 of the second acid were evaluated in each experiment, as well as the 16 mixtures of the 2 acids. Table 1 lists the concentrations used. The stimuli were prepared in 250-ml. quantities with distilled water (Hydro Service Supply, Inc.). At the time of preparation, the stimuli remained for 24 hr under refrigeration (4°C), following which they were frozen until use 2 wk. later. The solutions were thawed, remixed thoroughly, and served to *Ss* at room temperature (21°C ± 1).

Each stimulus was presented to *S* in small (½ oz.) paper cups containing 15-20 ml. of the solution. The stimuli in each set were presented in random order of concentration, compound, and mixture, and within any session, each *S* sampled all 24 stimuli (8 unmixed stimuli and 16 mixtures). The entire session required about 30 min., with *S* taking approximately 1 min. to sample the stimulus, record the sourness estimate, and then rinse with 50-100

TABLE 1

PARAMETERS OF SIMPLE SOURNESS FUNCTIONS

($S = kC^n$ AND $S' = k'C^{.7}$)

Acid	n	SE_n	log k	r^2	SE regression	log k'
Experiment I[a]						
Gluconolactone	.79	.10	1.36	.97	.09	1.34
Citric	.66	.09	1.66	.97	.09	1.73
Experiment II[b]						
Phytic	.88	.09	2.76	.98	.06	2.49
Citric	.62	.02	1.59	.99	.15	1.73
Experiment III[c]						
Succinic	.93	.09	2.17	.98	.07	1.83
Citric	1.01	.11	2.27	.97	.08	1.73
Experiment IV[d]						
Hydrochloric	.35	.03	.85	.98	.02	1.41
Citric	.92	.15	2.15	.95	.11	1.73

Note. Concentrations (molarity) were .05, .025, .0125, and .00625 for hydrochloric, succinic, citric, and phytic acids; and .80, .40, .20, and .10 for gluconolactone.

[a] $n = 8$.
[b] $n = 7$.
[c] $n = 6$.
[d] $n = 7$.

ml. of water provided for that purpose. The rinse water was room-temperature tap water.

All *S*s were military enlisted men (aged 18–24 yr.) who had had previous experience in judging the apparent sweetness of sugars by magnitude estimation. None were aware of the purpose of the study, and each had been previously trained on the use of magnitude estimation according to the procedure outlined by Moskowitz (1970). A pool of 10 *S*s participated, but smaller groups of individuals participated in any one session. The experiments consisted of one test in the morning and a replication of that test in the afternoon. The same *S*s participated in each session.

Analysis. The median magnitude estimate was computed for the 24 judgments for each set of stimuli. The median was selected as the preferred measure of central tendency because it is not as drastically affected by an erratic judgment as is the geometric mean. For each experiment, power functions in their linearized form ($\log S = n \operatorname{Log} C + \log k$) were computed by a least squares procedure, both for the median judgments of citric acid and for the median judgments of the second acid (in unmixed form). The goodness of fit was assessed by (*a*) the Pearson r^2, (*b*) the standard error of the regression, and (*c*) the standard error of the exponent.

The 2 theoretical models of sourness summation were used in conjunction with the simple power functions to predict the sourness of the 16 mixtures in each experiment. Each model uses both the exponent n and intercept k, according to Equations 4 and 5, respectively, and for every pair of acid concentrations, the 2 models predict different expected sourness values. In addition, estimates were made of the cosine of the angle separating the 2 "sourness vectors" for the vector model of additivity (Equation 6). Finally, in order to assess which model fits most appropriately the theoretical mixture sourness from Equations 4 and 5, theoretical mixture sourness (S_t) was related to the empirical estimate of mixture sourness (S_e) by the following equations:

$$\log (S_e) = n \log (S_t) + k_1, \qquad [7]$$

and

$$\log (S_e) = 1.0 \log (S_t) + k_2. \qquad [8]$$

Equation 7 states that the relation between predicted and obtained sourness is a power function. If the exponent n is 1.0, (Equation 8), then both predicted and obtained sourness differ only by a single value, k_2 (i.e., by a constant ratio). If the exponent n is higher than 1.0, then the mixture sourness grows more rapidly than the theoretically predicted sourness. As the sourness level increases, the mixture sourness will eventually be substantially higher than that predicted from the models. The opposite occurs if the exponent is lower than 1.0. Finally, if a single index number is desired to represent, on the average, how different the predicted and the empirical sourness values are, then Equation 8 is useful. The exponent is set arbitrarily to 1.0, and the least squares solution for k_2 provides that index number. If k_2 is positive, then the empirical sourness is stronger than that predicted from the additivity models; whereas if k_2 is negative, then the empirical sourness is weaker than that predicted. The former corresponds to "synergism," whereas the latter corresponds to "suppression."

RESULTS

Figure 1 shows the 4 pairs of sourness functions obtained from the median magnitude estimates of sourness, and Table 1 lists their respective parameters and the goodness-of-fit statistics. The acids are not described by the same power functions, but differ both in the exponent n and in the intercept k. The exponent of phytic acid, for example, is higher than that of citric acid judged in the same session. This means that equal percentage changes in concentration for the 2 acids produce a greater increment in the judged sourness of phytic acid than that of citric acid. Citric acid, in turn, is governed by a higher exponent than hydrochloric (a factor of almost 2.5:1), so that hydrochloric acid grows more slowly in sourness than citric acid for commensurate increments in molar concentration.

In the evaluation of simple, unmixed acids, the appropriate experimental procedure may be to determine sourness functions for both acids judged in the same session. In this way, the same *S* judges the 2 acids, using the same modulus, and

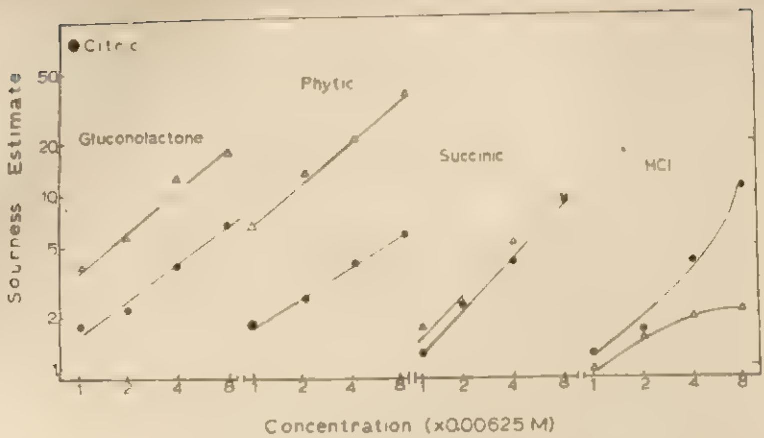


FIGURE 1. Sourness functions for 4 experiments. (In each experiment, citric acid and 1 other acid were each evaluated at 4 concentrations. The coordinates of the figure are log log, and straight lines represent power functions whose exponents are given by the slope of the line. The concentrations are in terms of molarity, and the median magnitude estimate is plotted against concentration.)

brings to bear in the experiment the same biases toward the operation of judgment. Were 2 acids to be evaluated in separate sessions by different *S*s, one *S* might use a larger range of numbers than another. The limited number range adopted by some *S*s can decrease the power function exponent considerably and has been called the "regression effect" in psychophysical judgment (Stevens & Greenbaum, 1966). If the regression bias is multiplicative or belongs to a class of transformations known as power transformations, then the relative exponents, determined under these conditions, would yield the same ratio, even though each exponent might be reduced or increased considerably. Here, the different *S*s may have influenced the variability of the exponent for citric acid from a low of .62 to a high of 1.01. This range of values lies within the range of exponents previously reported for various organic acids (Moskowitz, 1968, 1971).

In the evaluation of unmixed acid solutions, a single representative value for relative sourness is often desirable, if only

TABLE 2

PARAMETERS OF MIXTURE FUNCTIONS

Parameter	Mixture			
	Citric + gluconolactone	Citric + succinic	Citric + phytic	Citric + hydrochloric
Model 1: summation of concentrations				
n	.94(.17)[a]	1.04(.14)	1.09(.09)	1.18(.16)
k_1	.15	− .09	− .24	− .03
SE estimate	.14	.12	.10	.16
r^2	.68	.79	.90	.79
k_2	.10	− .06	− .11	.08
Model 2: summation of sournesses				
n	1.01(.14)	1.05(.15)	1.19(.11)	1.15(.16)
k_1	.02	− .11	− .48	− .01
SE estimate	.14	.12	.11	.15
r^2	.69	.79	.89	.79
k_2	.02	− .07	− .20	.09

Note. (*a*) log (Sourness$_{\text{empirical}}$) = n log (Sourness$_{\text{predicted}}$) + k_1. (*b*) log (Sourness$_{\text{empirical}}$) = 1.0 log (Sourness$_{\text{predicted}}$) + k_2.
[a] Numbers in parentheses refer to the standard error of estimate of n.

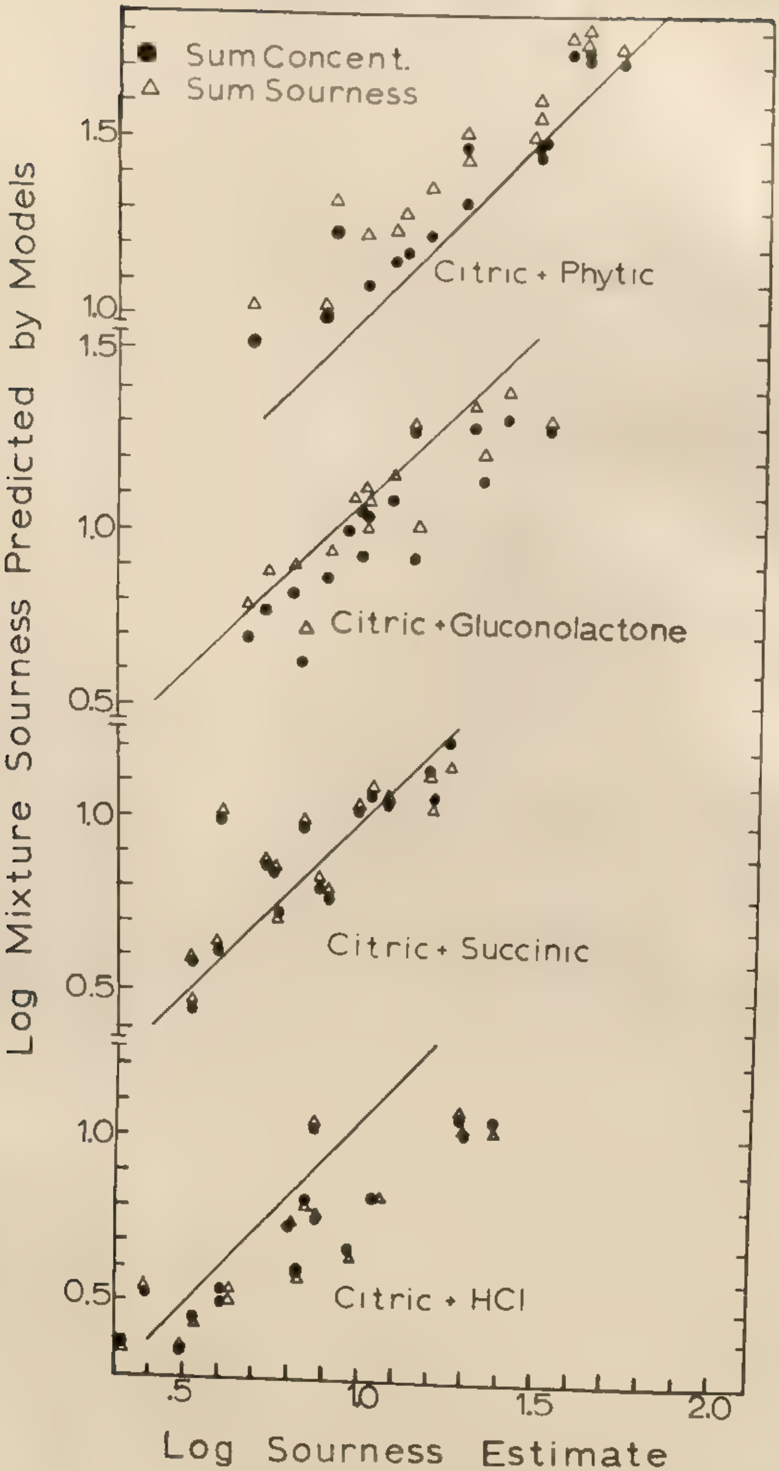


FIGURE 2. The relation between obtained (median) estimates of sourness for mixtures of 2 acids at varying concentrations, and the predicted sourness based upon 2 models of additivity. (The coordinates are log–log, so that

to serve as an index number for purposes of comparing different chemicals. The relative sourness for 2 acids may be immediately obtained by comparing 2 magnitude estimates of different acids (usually at 1 common concentration, e.g., 1.0 M). The ratio of estimates (corresponding to the difference in logarithmic values) corresponds to the desired ratio of sensory magnitudes. In contrast, the traditional (although misleading and ambiguous) definition of relative sourness has been the ratio of concentrations of 2 acids that produce an impression of the same sourness level. This latter ratio might better be called *taste potency* rather than relative taste intensity.

When exponents differ for acids, the relative sournesses tend to vary with concentration. In order to achieve a single index for the relative value, the 2 sourness functions must be forced to conform to a single power function with fixed slope (any slope or exponent suffices). The least squares estimate of the intercepts then serves as a useful value for sourness and can be used across the entire continuum of concentrations. In Table 1, the exponent has been set arbitrarily at .7 for each acid (independent of the actual least squares exponent) and the value k' corresponds to the intercept (at 1.0 M) for the .7 exponent. In all cases, the value for citric acid is 1.73 (in logarithmic units). Numbers lower than 1.73 represent lower sourness, so that .73 means that the second acid is 1.0 log unit lower in sourness (equal to 10% as sour as citric acid at equal concentrations). Gluconolactone has a sourness of 1.34, so that relative to citric acid, its sourness is $-.39$ (i.e., $1.34 - 1.73$). That value corresponds to approximately 40% relative sourness ($10^{-.39} = 40\%$). Phytic acid is the most sour and gluconolactone the least sour acid.

Acid mixtures. Table 2 presents the summary statistics for the analysis of acid mixtures, and Figure 2 presents the relation between the logarithms of the sourness estimates (empirically obtained judgments) of mixtures and the values for mixture sourness predicted by the summation models. The models were evaluated both by the summation of concentrations (expressed in equivalents of citric acid) and by the summation of sournesses (i.e., addition of power functions). In many instances, the mixture sourness exceeded the predicted sourness, and as a result, the vector model of additivity required a cosine greater than 1.0, an impossible case. Vector additivity could not account for synergistic effects, and was eliminated as a predictor model, based upon the relatively large number of synergistic mixtures.

The computations for mixture sourness in Table 2 and Figure 2 result from the composition of individual sourness functions (i.e., by adding together psychophysical power functions) rather than from the empirical sourness estimates themselves. As a result, the models are capable of accounting for mixtures of concentrations that are not directly tested in the experiment, since a function is used to transform concentration to sourness. In addition, however, those discrepancies between predicted and obtained sourness of mixtures may, in part, be traced to the choice of sourness functions for the unmixed acids. The increase in generality by appeal to a simple parsimonious function will be offset when that function provides only a modest fit to the sourness estimates of unmixed acids.

In Figure 2, the coordinates are logarithmic, and the best fitting functions in Table 2 refer to those coordinates. The straight line tilted 45° represents the line of congruity. When the predictions exactly

deviations from the 45° line—ideal congruence between predicted and obtained values—correspond to percentage deviations, not absolute deviations. Mixtures that are plotted above the 45° line represent pairs of acids whose predicted sourness is greater than the empirically judged sourness. Those mixtures show *suppression* or partial additivity. Mixtures plotted below the 45° line represent stimulus pairs whose predicted sourness is less than the empirical sourness. Those mixtures show *synergism*.)

equal the empirical sourness estimates, then all points fall along the line. Points *below* the line correspond to synergism, so that the obtained sourness estimates exceed the predicted values. Points *above* the line correspond to suppression, so that the obtained sourness estimates are less than the predicted values. Deviations from straight lines are logarithmic distances, corresponding to ratios in linear coordinates. The conversion to logarithms allows a constant percentage discrepancy to appear as a constant difference. In linear coordinates, that constancy would be difficult to ascertain by visual inspection alone. In addition, small deviations at the low end of the sourness scale are as important for the appropriateness of the models as are the larger deviations that would be expected at the higher end of the sourness scale. The small deviations would be masked substantially in linear coordinates.

A number of phenomena emerge from Figure 2. First, for the mixtures of citric acid with either gluconolactone or succinic acid, the mixture sourness grows as rapidly as that predicted from the mixture equations. The parallel functions do not appear for mixtures of citric with phytic and hydrochloric acids. Instead, when the sourness estimate is the dependent variable, the slope relating the 2 mixture sournesses exceeds 1.0, so that the 2 values depart increasingly from each other with increasing sourness. Second, departures from congruence, even when of a constant percentage nature (i.e., mixtures of citric acid with succinic acid and gluconolactone, respectively) can be either synergism (with gluconolactone) or slight suppression (with succinic acid). It is also interesting to note that phytic acid (which shows most suppression in mixtures) is the strongest of the acids in terms of the sourness it produces, whereas gluconolactone is the weakest. It is possible that relative sourness is implicated in synergism vs. suppression. Third, the 2 models produce highly correlated predictions that differ principally by a constant percentage. That percentage, rather than any other goodness-of-fit statistic, is the critical parameter by which the 2 models can be differentiated and evaluated. The percentage by which one model exceeds the other (Figure 2) is not constant, however, but varies systematically according to the component power functions that are used to generate the 2 mixture equations. For mixtures of citric and succinic acids, the models produce virtually the same predictions, whereas for mixtures with gluconolactone, the models predict moderately different sourness levels.

Discussion

The models of additivity suggested previously to account for the sweetness of sweetener mixtures (Cameron, 1947; Moskowitz, 1973) appear to be useful for predicting mixture sourness of acids as well. The present results bear upon 2 areas of taste psychology: (*a*) the analyzability of taste impressions and its correlation with taste intensity of mixtures, and (*b*) potential sensory mechanisms subserving taste impressions and their elucidation from studies of taste mixtures.

Analyzability. None of the *S*s in the present experiment appeared to be aware that they were sampling both simple stimuli and mixtures. Rather, the tastes were reported to be uniformly sour. Each *S* was questioned at the completion of the series of experiments, but none reported any other taste but sourness. Similar failure to perceive a taste as a mixture when both components possess the same quality was reported previously when fructose was added to glucose (Moskowitz, 1973). In that study, as in this, no stimuli contained bitter or other side tastes that might have interfered with sweetness or sourness, respectively. It may well turn out that like-tasting compounds add together according to a simple arithmetic manner, with synergism or suppression simply phemomena that reveal a failure to account in an adequate way for the law of summation.

In contrast to simple additivity of like-tasting compounds, mixtures with different tastes suppress each other. A study by Beebe-Center, Atkinson, Rogers, and O'Connell (1959) investigated the saltiness and sweetness of compound tastes containing NaCl and sucrose. The saltiness and sweetness were reduced by a total of 40% (in gusts, their unit of taste intensity), to 60% of the starting value. In a second study (Moskowitz, 1972), mixtures of sweet substances (glucose and fructose, 2

simple sugars) with NaCl (salty), citric acid (sour) or quinine sulfate (bitter) also showed suppresion, so that the total magnitude estimate given to the mixture (i.e., the sum of the magnitude estimates given to the 2 components) was less than the arithmetic sum of the estimates given to the unmixed components.

It is possible that the laws governing taste mixtures parallel *Ss*' ability to analyze the mixture into its components. Those mixtures in which analyzability is high (e.g., sweet + salty mixtures) may exhibit substantial suppression. Intermediate mixtures in which the components are qualitatively similar and merge slightly (e.g., some odor mixtures) show more additivity and less suppression. Finally, mixtures of like-tasting compounds may show total additivity, or at least a high degree thereof. It is also intriguing to note that an entire continuum of mixture phenomena is possible, ranging from synergism to complete (100%) additivity, through partial additivity, and down to suppression. Taste mixtures in general occupy 2 separate regions. Like-tasting compounds occupy the upper region, whereas different-tasting compounds occupy the lower region of suppression.

Underlying sensory mechanisms and taste mixtures. A study by Halpern (1959) compared the neural response at the medulla (nucleus fasisculus solitarius) to responses of the chorda tympani (peripheral taste nerve). At the chorda tympani homogeneous mixtures of inorganic salts produced an additive intensity response, so that the neural response to the mixture could be well approximated by summing the responses to the components. Heterogeneous mixtures of sucrose and citric or acetic acids also produced additive neural response (Halpern, 1967). In contrast, the central neural response at the medulla did not show a parallel summation. Instead, the magnitude was smaller than the expected sum, but greater than either response to the components separately. It would be instructive to compare central vs. peripheral neural responses to both homogeneous and heterogeneous mixtures and to then determine parallels between human psychophysical judgments of mixtures and neural responses.

REFERENCES

Beebe-Center, J. G. Standards for use of the gust scale. *Journal of Psychology*, 1949, **28**, 411–419.

Beebe-Center, J. G., Rogers, M. S., Atkinson, W. H., & O'Connell, D. N. Sweetness and saltiness of compound solutions of sucrose and NaCl as a function of concentration of solutes. *Journal of Experimental Psychology*, 1959, **57**, 231–234.

Berglund, B., Berglund, U., Lindvall, T., & Svensson, L. T. *A quantitative principle of perceived intensity summation in odor mixtures.* (Tech. Rep. 346) Stockholm: University of Stockholm Psychological Laboratories, 1972.

Cameron, A. T. *The taste sense and the relative sweetness of sugars and other sweet substances.* (Tech. Rep. 9) New York: Sugar Research Foundation, 1947.

Halpern, B. P. Gustatory responses in the medulla oblongata of the rat. Unpublished doctoral dissertation, Brown University, 1959.

Halpern, B. P. Some relationships between electrophysiology and behavior in taste. In M. R. Kare & O. Maller (Eds.), *The chemical senses and nutrition.* Baltimore: Johns Hopkins Press, 1967.

Moskowitz, H. R. Scales of intensity for single and compound tastes. Unpublished doctoral diseration, Harvard University, 1968.

Moskowitz, H. R. Ratio scales of sugar sweetness. *Perception & Psychophysics*, 1970, **7**, 315–320.

Moskowitz, H. R. Ratio scales of acid sourness. *Perception & Psychophysics*, 1971, **8**, 40–42.

Moskowitz, H. R. Perceptual changes in taste mixtures. *Perception & Psychophysics*, 1972, **11**, 257–262.

Moskowitz, H. R. Models of sweetness additivity. *Journal of Experimental Psychology*, 1973, **99**, 88–98.

Stevens, S. S. & Greenbaum, H. Regression effect in psychophysical judgment. *Perception & Psychophysics*, 1966, **1**, 439–445.

(Received July 13, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 648-656

A SEMANTIC INTERPRETATION OF ENCODING SPECIFICITY[1]

LYNNE M. REDER,[2] JOHN R. ANDERSON, AND ROBERT A. BJORK

University of Michigan

Two experiments are presented to clarify possible interpretations of the Encoding Specificity Principle of Tulving and Thomson. This principle states that a cue must have been studied with a word in order for the cue to be effective at testing. In the experiments reported here, recall and recognition of words were impaired by a change in the accompanying cues only if the to-be-remembered (TBR) words were of high frequency; low-frequency words did not support the Encoding Specificity Principle. The data suggest that both recall and recognition of a TBR word depend upon recognition of a specific interpretation of the word originally encoded, rather than its physical representation.

The considerable recent research on context effects in memory (e.g., Bobrow, 1970; Light & Carter-Sobell, 1970; Thomson, 1972; Thomson & Tulving, 1970; Tulving & Thomson, 1971; Winograd & Conn, 1971) has demonstrated that the ability to recall or recognize a word can be severely impaired by changes from study to test of the context in which the word is presented. Light and Carter-Sobell, for example, paired homographs with one adjective at the time of study (e.g., soda *cracker*) and later tested recognition of the homographs when paired with the same adjective or with an adjective that primed a different sense of the homograph (e.g., safe *cracker*). Recognition was significantly better when the same adjective was paired with the target word even though *S* knew that only the noun was to be judged for familiarity; thus, a word's context can greatly affect whether *S*s can access information stored with that word. This result is significant, since many models of memory (e.g., Anderson & Bower, 1972; Kintsch, 1970) assume that accessing a word's trace is an automatic process in tests of recognition memory.

Tulving and Thomson (1971) and Thomson (1972) have also found that recognition memory for to-be-remembered (TBR) words is impaired by changes in the cue word that is paired with a TBR word. Also, they found that adding or subtracting the cue word at the time of test impaired the recognition of TBR words. In contrast to the Light and Carter-Sobell (1970) procedure, however, Tulving and Thomson did not use transparent homographs, and they did not make a deliberate attempt to prime different meanings of the words. Thomson and Tulving (1970) also have found that cued recall of TBR words is lowered when the cue employed at the retrieval test is different from the encoding cue given at input, and that result holds even when the retrieval cue is a strong normative associate of the TBR word. Strongly associated, extralist cues (*strong cues*) enhance recall when no cues are presented with the TBR words at input (Bahrick, 1969; Thomson & Tulving, 1970), but are inferior probes compared to weak cues that were paired with TBR words at the time of input. It is possible in the Thomson and Tulving cued-recall situation that the *S* implicitly generates the TBR word to the strongly associated cue, but does not recognize the TBR word as being from the list. This interpretation is supported by the fact that *S*s can generate the TBR item in a free-association task to a strong cue, yet fail on a later

[1] This research was supported by the Air Force Office of Scientific Research under Contract F44620-72-C-0038 with the Human Performance Center, University of Michigan, and by a National Institute of Mental Health traineeship and a National Science Foundation graduate fellowship to the first author. We wish to thank Ed Martin for commenting on an earlier version of the manuscript.

[2] Requests for reprints should be sent to Lynne M. Reder, Human Performance Center, 330 Packard Road, Ann Arbor, Michigan 48104.

test to recognize that word as a TBR word (Tulving & Thomson, 1973).

Tulving and Thomson (1973) argue that only their Encoding Specificity Principle is consistent with these results. In defining this principle, they state that: "In its broadest form the principle asserts that only that can be retrieved that has been stored, and that how it can be retrieved depends on how it was stored [p. 359]." They assert that TBR words are stored as higher order episodic units and only elements of the *episodic* as opposed to *semantic* (see Tulving, 1972) higher order unit can access the TBR item. However, Light and Carter-Sobell (1970) reject the notion that their results could be due simply to "the effect of breaking up an encoded adjective-noun unit at recognition [p. 8]." They point out that words were recognized less frequently when a new adjective changed the sense of the noun than when the original adjective was simply deleted, and that recognition was better when the original adjective (e.g., soda *cracker*) was replaced by one that did not drastically alter the semantic interpretation of the target word (graham *cracker*) than when the new adjective did alter the semantic interpretation of the target word in a substantial way (safe *cracker*). Thus, Light and Carter-Sobell interpret their results as evidence that recognition of a word in a verbal learning experiment may involve recognition of the appropriate meaning of the word.

Work by Winograd and Conn (1971) supports the Light and Carter-Sobell (1970) interpretation. Words known to have multiple meanings were presented to *Ss* with no explicit context, that is, no encoding cue. Prior to the main experiment, the relative frequencies of the various interpretations of each word were determined. Recognition for these words was then tested by presenting a word in 1 of 3 ways: (*a*) In a sentence that used a frequent meaning or interpretation of the word; (*b*) in a sentence that used an infrequent meaning of the word; or (*c*) by itself, that is, with no context or meaning imposed by *E*. Homographs tested in a context that utilized their high-frequency meanings and those tested without an imposed context were recognized about equally well, and both were recognized far more often than the homographs forced into an uncommon encoding or meaning. Winograd and Conn concluded that words with multiple meanings are encoded with respect to a specific meaning even though no cue is presented at input and that, in the absence of an experimentally provided context, the most frequent meaning of a word tends to be the one encoded.

It might appear that the Light and Carter-Sobell (1970) and the Winograd and Conn (1971) analyses do not extend to the Tulving and Thomson data because Tulving and Thomson did not use explicit homographs. However, many words in the language that are not explicit homographs still have multiple senses or interpretations. For instance, *water* does not have 2 obviously different meanings, but the sense of *water* when presented with *lake* is somewhat different than the sense of *water* when presented with *drink*. In fact, *Webster's Seventh New Collegiate Dictionary* (1965) lists 8 main senses of the noun *water* and 16 subsenses and 19 sub-subsenses Included are senses unfamiliar to most people and some obscure senses—e g., "a capital stock not representing assets of the issuing company and not backed by earning power [p. 1006]"—that are unrelated to any of the normal senses of the word. It is also certainly the case that there are still other interpretations of the word that are not in the dictionary but are known to some *Ss*. Thus, many words not considered homographs allow multiple semantic interpretations, and a plausible explanation of the Tulving and Thomson data posits that a change in cue words can tap a different sense of the TBR word.

The experiments reported here were designed to test this semantic interpretation of encoding specificity. The basic idea is that words with few senses should be less vulnerable to changes in context in the

Tulving and Thomson paradigm than should words with many senses.

Although the idea behind this research is relatively straightforward, obtaining an adequate, objective measure of the number of senses associated with words is a nontrivial methodological problem. One cannot simply ask *S*s to report how many senses they have for particular words; it is not easy for *S*s to retrieve all the senses they have for a word (MacKay & Bever, 1967), nor is it easy to convince them that subtle distinctions like those associated with *water* are to be regarded as different senses. If it were easy to retrieve all the senses of a word, *S*s would not suffer the difficulties they do in experiments like those of Light and Carter-Sobell (1970).

For other reasons, number of dictionary meanings is also an inadequate measure. Many senses listed in a dictionary are not instantiated in most people's lexicon, while other colloquial and idiosyncratic senses are not in the dictionary. It is also difficult in practice to be consistent in terms of the cutoff point one uses to decide that a dictionary distinction is in fact substantial enough to denote separate senses of a word.

Based on these and other considerations, word frequency was chosen as the index of number of word senses. Words of very low frequency (less than 9/1,000,000 in Kučera & Francis, 1967) were contrasted with words of moderate to high frequency (more than 50/1,000,000). Generally, low-frequency words (e.g., hippopotamus, aspirin) tend to have only one sense. This phenomenon is probably attributable to there being little opportunity for multiple senses to differentiate and be maintained. It may also reflect the fact that words with one unique sense are appropriate to only a few situations and as a result have low frequency of occurrence. Therefore, word frequency is a fairly sensitive correlate of number of senses instantiated in the typical *S*. Schnorr and Atkinson (1970) found a strong positive correlation between word frequency and number of dictionary meanings, as did we for the words used in the experiments reported here, $r = .61$; $t\ (118) = 8.3$, $p < .001$.[3] Both word frequency and number of dictionary meanings are undoubtedly imperfect reflectors of number of senses, but frequency seems to provide a somewhat more satisfactory and objective basis on which to partition words into those having few or many senses. Later in this report frequency and number of dictionary meanings are compared in terms of their ability to predict memory performance.

Experiment I

Method

Design and procedure. Four lists of 30 TBR words, 15 high frequency and 15 low frequency, were presented by slide projector at a 3.5-sec. rate to *S*s for study and subsequent recall. Each TBR word was initially paired with a weakly associated cue; the TBR word was shown in uppercase letters directly below the cue word, which was shown in lowercase letters. After presentation of 30 pairs, 10 of the TBR words were tested in each of the 3 following conditions: (*a*) The *S* saw the original weakly associated input cue and was asked to recall the TBR word that appeared with it; (*b*) *S* saw a new cue strongly associated with a TBR word and was asked to use it to help him think of a TBR word from the last list; or (*c*) *S* received no cue for the word and was simply asked to try to recall the item from memory. Within each list, the 2 levels of word frequency were crossed with the 3 types of output conditions. There were 5 words in each Frequency × Output combination. Three different sets of recall sheets were used for each list so that, across *S*s, all words appeared in all output conditions. Each set consisted of 2 recall sheets, one for cued recall and one for the recall of words for which no probe or cue was given. The sheet for cued recall was presented before the free-recall sheet for half of the *S*s and after the free-recall for the other *S*s. The *S*s had as much time as they wanted for both uncued and cued recall. When uncued recall preceded cued recall, *S*s were asked to recall all the TBR words; when uncued recall followed cued recall, *S*s were told that they only needed to recall those words from the list that they had not recalled on the cued-recall sheet, but that they need not worry about duplications.

The word presentation order within a list and the order of the cue words on the cued-recall

[3] We had a group of *S*s rate all the words in our experiments for number of different meanings they could think of for each word. The *S*s were only able to generate a mean of 1.93 meanings per word whereas Webster's dictionary has 8.83 meanings. The number of generated meanings correlates .64 with frequency and .67 with number of dictionary meanings.

sheets were randomly determined but were held constant across *S*s. Half of the 20 cues were asterisked to indicate that they were new cue words. The various recall sheets and the order of list presentation were counterbalanced over *S*s. Two priming lists, which were not scored, preceded the 4 lists of interest. The priming lists, composed of medium-frequency words, were tested at recall with their original input cues. The purpose of the priming lists was to induce *S*s to encode the TBR word with respect to the original cue presented; *S*s were not informed that the priming lists were practice lists. Forty-eight paid *S*s at the University of Michigan were tested in groups of 1–7.

Materials. All TBR words were nouns judged to be high in both meaningfulness and imagery. Items were selected from norms of frequency of words in print (Kučera & Francis, 1967); the high-frequency words came from the upper third of the rank ordering in frequency, and the low-frequency words came from the bottom third; words in the priming lists were taken from the middle range. (The high-frequency words all had A or AA Thorndike & Lorge, 1944, ratings, and the low-frequency words averaged 6.7/1,000,000 in the Thorndike and Lorge count.) Those nouns selected that were included in Paivio, Yuille, and Madigan's (1968) rating of meaningfulness and imagery of nouns received scores that confirmed the subjective criterion for high imagery and concreteness: They all were above 6 on the two 7-point scales. Thomson and Tulving (1970) constructed their lists of TBR words by selecting response words from free-association norms in which each TBR word occurred twice in the norms: once as a high-frequency response to a *strong cue* stimulus word and once as a low-frequency response to a different *weak cue* stimulus word. Association norms could not be used to select the cues for this experiment due to the difficulty of finding norms that contain low-frequency nouns as responses to other words, particularly as strong associates.

Because of these inherent difficulties, strong and weak cues were generated according to intuitive criteria. That is, the strong cues seemed likely to elicit the TBR words frequently, the weak cues infrequently. The fact that the cues were constructed in an informal manner does not seem a serious problem in methodology. The basic requirement is simply that the strong cues for low-frequency words are not more effective than strong cues for high-frequency words. This requirement is important because the predicted results (namely, that recall of low-frequency words with their new, strong cues will be superior to recall of high-frequency words with their new, strong cues) might otherwise be an artifact of prior strength. Five colleagues, naive with respect to the study, gave free associations to each of the strong cues used in the experiment. High-frequency TBR words were generated as free associates to the strong cues 24% of the time; low-frequency TBR words were generated only 12% of the time.

TABLE 1

Proportions of High- and Low-Frequency To-Be-Remembered Words Recalled on Tests of Cued and Free Recall

Word frequency	Cued recall		Free recall
	Strong extralist cue	Weak within-lists cue	
Cued recall before free recall			
High	.31	.69	.04
Low	.64	.72	.05
Cued recall after free recall			
High	.36	.63	.12
Low	.59	.69	.14

The other cue-construction requirement used by Thomson and Tulving (1970) is that the strong and weak cues for a given TBR word should not be associatively related. This requirement was impossible to maintain with low-frequency words. For 2 words to be both associatively related to a third word and yet not be associatively related themselves tends to require that the third word have more than one interpretation, and low-frequency words often do not. The only requirement in the current experiment was that no 2 cues to a given word be synonyms.

Results and Discussion

In Table 1, recall probabilities are shown as a function of type of test, order of tests, and TBR word frequency. The statistical analyses reported in this section were carried out on arc sine transformations of the proportions in Table 1.

Cued recall. Since performance on tests of cued recall was not sensitive to testing order, F (1, 46) = 1.2, the other statistical analyses of the cued-recall data were carried out on the pooled data. The result of greatest interest in this experiment, that is, the frequency by cue (strong vs. weak) interaction, was highly significant, F (1, 46) = 18.3, $p < .001$, with performance in the high-frequency strong-cue condition much worse than performance in the other conditions. For high-frequency words, strong extralist associates were much less effective as recall cues than were weak intralist associates, F (1, 46) = 77.44, $p < .001$. In the case of low-frequency words, there is a smaller but still signifi-

cant advantage in the same direction, $F(1, 46) = 7.56$, $p < .01$. The latter result is not too surprising since low-frequency words are sometimes polysemous, and in such cases a change of cue would have adverse effects.

In order to rule out the possibility that the foregoing results might be attributable, in part, to item selection (cf. Clark, 1973), the data were also analyzed by collapsing over *S*s and treating words as the random effect. The Critical Frequency × Cue interaction was still highly significant, $F(1, 118) = 13.51$, $p < .001$. Even with both significant *F*s, however, the results may still be colored by *S*–item selection. Clark prescribes the use of a *min F′* statistic, $(F_1F_2)/(F_1 + F_2)$, a conservative test, the significance of which guarantees generalizability of a result over *S*s and items. The *min F′* statistic is significant for both the Frequency × Cue interaction, $F'(1, 149) = 7.77$, $p < .01$, and for the effect of weak vs. strong cues for high-frequency words, $F'(1, 135) = 37.91$, $p < .001$. On the basis of this more conservative test, however, the effect of weak vs. strong cues was not significant for low-frequency words, $F'(1, 142) = 3.45$.

Free recall. The results of the free-recall test in Experiment I are relatively straightforward. Overall, *S*s were a great deal less likely to free recall the 10 TBR words not tested on the cued-recall sheets than they were to recall TBR words in response to either type of cue. In contrast to the cued-recall performance, there was a distinct effect of test order, $F(1, 46) = 41.9$, $p < .001$. The test of cued recall appears to have provided substantial interference with subsequent free recall of TBR words not tested during the cued recall, whereas cued recall was independent of whether there was or was not a prior free recall.

Experiment II

Our interpretation of the context effects on cued recall in Experiment I involves a generation–recognition model for recall (see Anderson & Bower, 1972; Bahrick, 1969; Kintsch, 1970). However, the model we have in mind, unlike the prior generation–recognition models, assumes that *S* does not recognize the words per se, but rather, semantic interpretations of these words. The *S* implicitly generates words and attempts to recognize their senses. If he recognizes the senses as being from the TBR list, then he recalls the corresponding word. The difficulty caused by a change in context is attributable to the recognition phase of recall; *S* can generate the TBR word to the strong cue but he cannot recognize it because he assigns a different interpretation to it than he assigned when the word was paired with the weak cue. In Experiment II, *S*s were instructed to generate free associations to strong cues, and then to recognize if any of the associates were TBR words. If the difficulty of recalling high-frequency words to new, strong cues results from difficulties in generation, then new, strong cues should have a greater propensity to elicit their low-frequency TBR words than would the corresponding new, strong cues to their high-frequency TBR words. However, we expected this not to be the case. Rather, we expected that high-frequency TBR words, when generated to the cues, would not be recognized as well as low-frequency TBR words.

Method

Subjects. Thirty undergraduates at the University of Michigan participated as paid *S*s. The *S*s were tested in groups of approximately 5.

Materials and apparatus. Four of the 6 lists used in Experiment I were used in Experiment II—the same 2 priming lists and 2 of the 4 critical lists. Each list consisted of 30 TBR words, again the priming lists being all medium-frequency words, while the critical lists were half high-frequency words and half low-frequency words. All words in each list were paired with a weak cue. Each list was presented by means of a Kodak Carousel slide projector.

Procedure and design. The 2 priming lists preceded the 2 critical lists. The procedure followed and the reason for including the priming lists in the design were the same as in Experiment I.

After presentation of the first of the 2 critical lists, there was the following sequence of events.

1. The *S*s were first given a sheet of paper with 30 strong extralist cues listed in a column on the left-hand side of the sheet. They were asked to generate exactly 4 free associates to each strong cue. It was pointed out to *S*s that their

free associates to a given cue might contain a TBR word from the preceding list. They were asked to circle any such words they generated.

2. After the free-association and recognition tasks, *Ss* were given a 4-alternative forced-choice (4-AFC) recognition test consisting of 30 sets of 4 words, each set containing 1 of the 30 TBR words presented in the first critical list. The 3 distractor words in each set were drawn from words generated by several individuals asked to free associate to word pairs consisting of a TBR word and its strong cue. The distractors were selected to have about the same frequency as the TBR word.

3. Finally, in order to motivate *Ss* to attend to the weak within-list cues presented in the second critical list, there was a test of cued recall involving the weak within-list cues used in the first critical list.

The presentation of the second critical list was followed by the same sequence of events, except that when *Ss* free associated, they were asked to think of as many meanings as possible for each word generated. That is, *Ss* were asked to generate 4 words to the cue, but then to try to examine each of these words for meanings not related to the cue that elicited them. This instructional manipulation was motivated by the idea that it might reduce the inhibiting effect of the semantic context induced by the strong cue on the recognition of TBR words generated by *Ss*.

Results

The data from Experiment II are presented in Table 2. The effect of instructions, i.e., to think of multiple senses of the generated words during the free-association task, was insignificant with respect to all 3 dependent variables so the data are pooled over the 2 critical lists.

As expected, the probability of generating a high-frequency word to the new, strong cue was not lower than that for a low-frequency word; if anything the reverse was true: High-frequency TBR words were generated with probability .59 to their respective cues, and low-frequency TBR words were generated with probability .55 to their strong cues. The probability that a low-frequency TBR word was recognized (.84), however, given that it was generated during the free-association task, was more than twice the probability that a high-frequency word was recognized (.38) given it was generated. The probability of recognizing low-frequency TBR words in the 4-AFC

TABLE 2

Recognition and Cued Recall Probabilities as a Function of Test Type and Word Frequency

Word frequency	Type of test: Generation-recognition	4-AFC recognition	Cued recall
High	.38 (.59)	.72	.61
Low	.84 (.55)	.85	.71

Note. The values in parentheses are the probabilities that a to-be-remembered word was generated in the free-association task. Abbreviation: 4-AFC = 4-alternative forced-choice recognition test.

recognition test (.85) was also higher than the recognition of high-frequency words (.72), although the difference was much smaller.

The recognition proportions for individual *Ss* were converted via the arc sine transformation and the transformed scores were subjected to an analysis of variance with the independent variables being word frequency and method of test. The Frequency × Test Type interaction was highly significant, $F\ (1, 29) = 30.34$, $p < .001$, as was the difference between recognition of *S*-generated high- and low-frequency TBR words, $F\ (1, 29) = 117.56$, $p < .001$, and the difference between the recognition of high- and low-frequency TBR words on the 4-AFC test was also significant, $F\ (1, 29) = 9.39$, $p < .01$. Again, in consideration of possible selection artifacts, *min F′*s were computed for the Frequency × Type of Recognition Test interaction, $F'\ (1, 73) = 16.25$, $p < .001$, the difference between recognition of high- and low-frequency generated TBR words, $F'\ (1, 87) = 41.93$, $p < .001$, and the difference between frequency levels in the 4-AFC, $F'\ (1, 50) = 7.02$, $p < .025$.

The fact that the discrepancy between the recognition of high- and low-frequency TBR words is less on the 4-AFC test than it is for the recognition of *S*-generated TBR words is probably attributable to differential effectiveness of the context induced by the test. The meanings of *S*-generated TBR words are clearly determined by the strong cue from which they are generated. However, in the 4-AFC test, *Ss* seemed better able to judge

a word without regard to the other alternatives. Since the strong cue was not actually present on the 4-AFC test, it was, presumably, easier to think of the word with a meaning different from that imposed by the strong cue and more like that imposed by the original cue. Tulving and Thomson (1973) also found improved recognition of TBR words when the alternatives in the 4-AFC test were not generated by the S[4] to the strong extralist cue.

Recall to the original cues yielded results very similar to the corresponding conditions in Experiment I: The cued recall probability for high-frequency words was .61 and the corresponding probability for low-frequency words was .71.

General Discussion

The present results necessitate some refinement and clarification of Tulving and Thomson's (1973) Encoding Specificity Principle. It seems plausible, at least for retention intervals on the order of those in the present paradigm, that what is recognized in a recognition test is a particular sense of a word rather than phonemic or orthographic information. A cue may generate the same word presented earlier, but if the semantic interpretation imposed by the cue in this generation process is different from the meaning originally encoded, the word is unlikely to be recognized or recalled. It may be that if the word is tested after a brief retention interval, then phonemic cues will improve recognition. However, in experiments such as these in which Ss must use long-term memory at testing, the information retrieved would probably be a sense of the word, not its spelling pattern or phonemic characteristics per se. Past studies have shown that long-term memory confusions are semantic in nature rather than acoustic as they are in short-term memory (e.g., Craik, 1968). Furthermore, since the relation between a TBR word and its cue is a semantic one, it is even more likely in cuing experiments than in other experiments that an S would tend to the meaning of the TBR word.

Tulving and Thomson (1973) assert that the generation–recognition models for recall (Anderson & Bower, 1972; Bahrick, 1969; Kintsch, 1968, 1970) are incompatible with the findings that support their Encoding Specificity Principle. They argue that generating a word does not guarantee that it will be recognized (which is the requirement for recall) even though the "event information," as they call it, is *available*. Previous descriptions of the generation–recognition model (e.g., Anderson & Bower, 1972) stipulated that if the TBR item was encoded as occurring in the list, only implicit generation of the appropriate spelling pattern (or phonemic cues) was required for recall. Tulving and Thomson showed that this was not the case. However, if the generation–recognition models are modified by the assumption that word senses are the basis of recognition, then these models become compatible with their principle. Words, per se, are not generated to a cue, but rather senses of a word are generated. The generated meaning must match a meaning encoded during the experimental task before recall can occur.

Effect of Strong versus Weak Cues

We proposed that many words have multiple senses and frequency of the word in print is a good index of how many senses words have for a particular S. Some senses of a word will unavoidably be used and encountered more often than others. A strong cue, as manifested by its high propensity to elicit the TBR words, probably selects a salient sense of the TBR word. On the other hand, weak associations tend to be bizzarre and thus tap less likely interpretations of the TBR words. When S sees a strong cue and generates the TBR word, the sense or interpretation which comes to mind is totally different from the obscure one studied, and S frequently cannot think of all the possible meanings related to a word generated and thus does not "notice" the studied sense. Therefore, it would seem that the more obscure or removed the old interpretation, the less likely that it will be noticed, and consequently, the less likely the word would be recognized.

Tulving and Thomson noted that their theory could not account for the finding that weak input cues paired with strong output

[4] The difference between recognition of words generated by S and recognition of words that were generated by a yoked S was in the same direction as our results although the difference was smaller. The smaller difference may be due to the fact that in their experiment the strong cue words were presented along with the distractors.

cues produced higher recall than strong input cues and weak output cues, as they expected no difference (Thomson & Tulving, 1970). If one assumes, however, that recall requires both generation and recognition, then one should expect better recall when the output cues are strongly associated rather than weakly associated because, in the task of recall, the candidates must be thought of before they can be recognized as words from the list. There is simply a much higher probability of *generating* the TBR word with the strong cue, regardless of the 2 cues' relative effects on recognition. Therefore, one would have to predict that weak input cues with strong recall cues give better performance than the converse.

Given the generation–recognition distinction and the assumption that it is senses that are recognized, the predictions for recognition are the opposite of those made for recall: A TBR word presented with a strong cue and later tested with a weak cue should be better recognized than a word originally presented with a weak cue and later tested with a strong cue. This prediction follows because the TBR word in the presence of a strong associate is not likely to suggest the obscure alternate sense of the word studied during its presentation with a weak cue. On the other hand, if the TBR word were originally presented with a strong cue, a frequent interpretation of the word would be studied. Because this sense is so salient, *S* would be likely to revive that same sense even when the word is presented in the context of a new, weak associate. This prediction was also confirmed in Tulving and Thomson's (1971) data, but no explanation was offered for that outcome.

Thomson (1972) also did not explain a result from one of his recognition experiments in which context was deleted. He found that presenting only 1 of 2 words originally presented as a pair had a less deleterious effect when the original pair had a strong associative relation than when it had a weak one. Also, the most harmful effect on recognition performance occurred on tests of single words initially presented on the right side of weakly associated input pairs. One possible explanation is that strongly related pairs tend to be encoded with the common sense of each word, so if either of the words is presented separately, the original sense is very likely to be noticed. On the other hand, when 2 words are only weakly related, at least 1 of the 2 words must be encoded in terms of a less frequently used sense of the word. It seems reasonable that the first word encountered when reading the pair imposes its dominant sense on the second word (unless, for some reason, a more obvious interpretation employs a common sense of the second and an obscure sense of the first). Thus, it seems plausible that a word on the right in a weakly associated pair would be likely to have an infrequent and unusual meaning encoded, which, in turn, would produce poor subsequent recognition of the word.

Effects of Word Frequency

As stated earlier, we think that word frequency is a better index of the average number of senses of a given word than is number of dictionary meanings. Although the 2 indices are strongly correlated for the words in this experiment ($r = .61$), the frequency measure correlated better with the obtained results. Using the data of Experiment I, a difference score was computed for each TBR word by taking the number of correct recalls to the original weak cue and subtracting the number of correct recalls to the new, strong cue. This difference score correlated .37 with frequency and .17 with number of dictionary meanings. The correlation between number of dictionary meanings and the difference score when frequency was partialed out was $-.06$ and frequency correlated .33 with the difference score when the number of dictionary meanings was partialed out. The first partial correlation is negative and not significant, $t\ (117) = .65$, $p < .60$. The second is significant, $t\ (117) = 3.78$, $p < .001$.

Rubenstein, Garfield, and Millikan (1970) provide incidental support for the notion that high-frequency words are basically quite similar to homographs and that frequency may be a better indicator of number of senses in a person's lexicon than dictionary entries. They administered a task that showed that decisions concerning whether letter strings are words had shorter latencies for homographs than nonhomographs. Their explanation was that homographs have more lexical entries and hence the probability of matching the stimulus with an entry is greater in any period of time, thus making reaction times shorter for homographs. Rubenstein et al. also found a highly significant effect for frequency of the word, high-frequency words being faster than lower frequency words; in fact, in a reanalysis of their data, Clark (1973)

found that the homography effect per se was insignificant, but that the strong frequency effect remained. Neither Rubenstein et al. nor Clark considered the notion that frequency might be a very sensitive measure of the number of meanings for a given word in a particular *S*.

REFERENCES

Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, **79**, 97–123.

Bahrick, H. P. Measurement of memory by prompted recall. *Journal of Experimental Psychology*, 1969, **79**, 213–219.

Bobrow, S. A. Memory for words in sentences. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 363–372.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language sampling in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–359.

Craik, F. I. M. Types of error in free recall. *Psychonomic Science*, 1968, **10**, 353–354

Kintsch, W. Recognition and free recall of organized lists. *Journal of Experimental Psychology*, 1968, **78**, 481–487.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Kučera, H., & Francis, W. N. *Computational analysis of present-day American English*. Providence: Brown University Press, 1967.

Light, L. L., & Carter-Sobell, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 1–11.

MacKay, D. G., & Bever, T. A. In search of ambiguity. *Perception & Psychophysics*, 1967, **2**, 193–200.

Paivio, A., Yuille, J. C., & Madiga[illegible] S. A. Concreteness, imagery, and meaningfuln[illegible] values for 925 nouns. *Journal of Experimen[illegible] Psychology Monograph Supplement*, 1968, **76**(1[illegible]t. 2).

Rubenstein, H., Garfield, L., & Millika[illegible] J. A. Homographic entries in the internal [illegible]icon. *Journal of Verbal Learning and Verbal B[illegible]ior*, 1970, **9**, 487–494.

Schnorr, J. A., & Atkinson, R. C. Stud[illegible] position and item differences in the short- an[illegible]ong-term retention of paired associates lear[illegible] by imagery. *Journal of Verbal Learning and [illegible]rbal Behavior*, 1970, **9**, 614–622.

Thomson, D. M. Context effects in reco[illegible]tion memory. *Journal of Verbal Learning and [illegible]bal Behavior*, 1972, **11**, 497–511.

Thomson, D. M., & Tulving, E. Asso[illegible]tive encoding and retrieval: Weak and stron[illegible]ues. *Journal of Experimental Psychology*, 197[illegible] **86**, 255–262.

Thorndike, E. L., & Lorge, I. *The te[illegible]her's wordbook of 30,000 words*. New York: C[illegible]mbia University, Teachers College, Bureau of [illegible]lications, 1944.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organisation of memory*. New York: Academic Press, 1972.

Tulving, E., & Thomson, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Expe[illegible]ogy*, 1971, **87**, 116–124.

Tulving, E., & Thomson, D. M. [illegible] specificity and retrieval processes in [illegible] memory. *Psychological Review*, 1973, **80**, [illegible]73.

Webster's seventh new collegiate dictionary. (7th [illegible]d.) Springfield, Mass.: G. & C. Merriam, 1965.

Winograd, E., & Conn, C. P. Evidence from recognition memory for specific encoding of unmodified homographs. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 702–706.

(Received July 14, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 657–669

RECALL AND RECOGNITION OF PICTURES BY CHILDREN AS A FUNCTION OF ORGANIZATION AND DISTRACTOR SIMILARITY[1]

JEAN M. MANDLER[2]
University of California, San Diego

NANCY L. STEIN
Washington University

Verbal description, recall, and recognition of complex meaningful pictures by children were studied, varying amount of organization in the stimuli and similarity of distractors. Across *S*s (sex, ethnic group, and grade level), verbal measures were poor predictors of recognition accuracy. Across stimuli, amount recalled and recognition accuracy were both related to amount of organization. Recognition was also a function of the type of transformations on the target that were used as distractors. For all transformations there was a close match between ability to recognize a transformation and judgment of dissimilarity of the transformation to the target. The TORSCA multidimensional scaling technique was applied to the similarity judgments to obtain a representation of a memory space for the targets and their transformations. The structure of this space was highly consistent across *S*s and indicated that transformations on meaningful pictures can be related to each other in stable ways. Location of transformations within the space was related to type of picture and amount of organization in the picture.

Investigators exploring the visual storage systems of both adults and children (Brown & Scott, 1971; Shepard, 1967; Standing, Conezio, & Haber, 1970) have shown that once *S*s are exposed to pictorial material, their ability to recognize previously seen pictures is extremely good. Standing et al. presented 2,560 pictures and found a hit rate for old pictures of .95 or better. Haber (1970) suggested that picture recognition may be essentially unlimited, and has contrasted it with recognition for verbal materials. "Unlimited" is perhaps too strong a term, however, since several variables are known to affect accuracy of picture recognition.

1. Similarity of distractors to target: Recognition of pictures of single objects is directly related to similarity of the distractors, when the latter are scaled along a single "similarity" dimension (Bahrick, Clark, & Bahrick, 1967). When more complex pictures have been studied, however, few experiments have systematically manipulated this variable. Most studies have used a random selection of distractors, chosen from the same modality as the target items (e.g., photographs) but with content uncontrolled. One study which used similar distractors (Dallett, Wilcox, & D'Andrea, 1968) found somewhat lower recognition rates, ranging .69–.85, but no attempt was made to define or scale degree of similarity. Brown and Campione (1972), working with children, studied a single dimension of similarity by using distractors that varied from the targets only in the type of action a central character performed. Immediate recognition averaged around 95%. Thus, for adults, recognition accuracy is affected by similarity of distractors, but there are not enough data to know whether this is as true for children. In addition, little is known about the dimensions along which children scale similarity of remembered pictures to distractor items.

2. Organization of the stimulus materials: Organization of a picture is defined here

[1] This research was supported in part by a grant from the San Diego Unified School District to the authors, and in part by U.S. Public Health Service Grant MH-15828 to the Center for Human Information Processing, University of California, San Diego. Responsibility for authorship is equally shared. We would like to thank George Mandler and participants in a research seminar at University of California, San Diego, for many creative suggestions during the course of this research.

[2] Requests for reprints should be sent to Jean M. Mandler, Department of Psychology, University of California, San Diego, La Jolla, California 92037.

by the relations among its various parts and is equated with the concept of meaning (see Garner, 1962, p. 141). Adults recognize pictures of faces better than inkblots (Goldstein & Chance, 1971), and children recognize meaningful pictures better than abstract or nonsense pictures (Nelson, 1971). However, little is known about the effects of higher levels of organization on recognition, namely, the relationships within a picture among parts that are themselves meaningful, such as are usually found in complex pictures.

It seems likely that judgments of similarity will vary with the amount of organization (i.e., number of meaningful relations among objects) in a picture. There is little information on this problem, yet it seems reasonable to assume that the more information that is organized into a single chunk (or picture) the less likely it is to be confused with another chunk containing an equal amount of different information. Thus, if complex pictures do contain a very large amount of both visual and semantic information, it is probably the case that the pictorial distractors used in experiments such as that of Standing et al. (1970) were grossly dissimilar from the target items. This brings us to the third variable known to affect accuracy of picture recognition.

3. Verbal or semantic information: Several studies have assumed that using pictorial distractors that have the same verbal label as the target pictures will increase distractor similarity and therefore decrease recognition. In 2 studies with children this was not the case; accuracy increased under this condition (Rosinski, 1970).[3] A decrease in recognition accuracy has been reported with adults, however (Bahrick & Boucher, 1968). Further, verbal labels attached to a geometric form have been shown to produce systematic changes in recognition accuracy along a continuum of physical similarity (Daniel, 1972). Clearly, one of the primary problems facing the investigator who is interested in recognition of meaningful pictorial materials by children or adults is the relation of purely visual dimensions, such as size or shape, to the semantic dimensions of pictures. For example, we do not know whether changes in orientation, deletions, or substitutions, affect recognition of meaningful collections of objects in picture form in the same way as they do for geometric shapes.

The present experiment investigated several of these aspects of picture recognition in children. The degree of organization of complex pictures was va[illegible]s well as the similarity of distracto[illegible] Distractors were constructed by [illegible]ing transformations on the original pictures. Two recognition tests were used, one measuring recognition accuracy and th[illegible]r requiring judgments of similarity of t[illegible]ious transformations to the remem[illegible] pictures. Multidimensional scaling techniques were used to construct dimensions of similarity of the various transforma[illegible] to the originals. In addition, verbal description and verbal recall of the pictures were studied.

One further problem was investigated in this experiment. It has frequently been suggested in the literature that black children may have more difficulty in handling pictorial material than white children (Farnham-Diggory, 1970; Pettigrew, 1964; Tyler, 1956). In spite of its widespread currency the hypothesis has never been adequately tested. Most of the studies used to support the hypothesis have contrasted performance on visual and verbal tasks in IQ test batteries. Such test batteries are not well-suited for measuring perceptual functioning, since each subtest usually involves a number of cognitive factors (e.g., Cohen, 1959). The present experiment provided an opportunity to test the hypothesis more directly, since 2 aspects of recognition of pictorial materials were studied, using measures relatively uncontaminated by verbal requirements. For this reason, as well as to insure generality of results, children from black, Chicano, and white (Anglo) ethnic groups were used as *S*s.

Method

Subjects

Because extensive verbal and recognition data were collected, a large number of *S*s was used, each working with only 2 target pictures, rather than a

[3] The second study is an unpublished report by L. R. Brooks written in 1971.

FIGURE 1. Organized and unorganized versions of 2 pictures.

smaller number of *S*s, each viewing many pictures. The *S*s were 168 children from the San Diego Unified School District, divided into groups by sex, ethnic group (black, Chicano, and white), and grade level (Grades 1 and 2). Chronological age in the first grade ranged 6.2–7.1 yr. ($M = 6.75$) and in the second grade, 7.2–8.3 yr. ($M = 7.8$). Almost the entire first and second grade English-speaking classes of one school, located in an upper-lower socioeconomic status area, were used as *S*s. The area was composed of a fairly even mix of the 3 ethnic groups. United States Census Tract data indicated a comparable socioeconomic status for the 3 groups in terms of employment. Median family income in the area was $7,075, with Chicanos having slightly higher income ($7,415) than whites, and blacks slightly lower incomes ($6,680). Median amount of schooling for adults in the area was 11.9 yr. Black adults were above the median (12.3) and Chicano adults slightly below (10.6 yr.). There were not enough black children to complete the sample, and slightly less than half of the second-grade black sample was drawn from another school from a somewhat higher socioeconomic status area, with a higher median family income of $10,150, but the same median schooling for adults of 12.4 yr.

Stimuli

Two sets of 21.6 × 27.9 cm. black-and-white line drawings were used; each set contained one "organized" and one "unorganized" picture. Within a set, the organized and unorganized pictures contained the same objects. In the organized version, a naturalistic scene was created, using familiar relationships among objects. In the unorganized version, the same objects were placed in an unrelated array. Because of some overlapping of objects in the organized pictures, there was more empty space at the edge of the pictures than in the unorganized versions. In the latter pictures there was more space between the objects. It was hoped that this difference would emphasize differences in processing of relationships among objects in the 2 types of picture, but it should be noted that the relationship of this variable to pictorial organization in general has not yet been systematically explored. The 4 pictures are shown in Figure 1.

The pictures were constructed during pilot testing. Easily recognizable objects with a considerable amount of detail were drawn by 2 amateur artists. Final selection was made on the basis that the various objects could fit into a "realistic" scene and that in the pilot group children from the same ages and ethnic groups used in the study could appropriately label all objects and details. A rough attempt was made to equate number of objects in the 2 sets of pictures. However, it is recognized that the pictures vary in a number of ways, such as saliency and location of items, likely focus of attention, etc.

The following 5 transformations were used for each of the pictures:

1. Reversal: A left-right mirror image of the original.

2. Deletion: Three items were deleted from the original, ranging in size from a small detail to a small object. In the bus pictures (see Figure 1), the girl's purse, the chimney on one house, and one of the books were deleted. In the lady pictures, the pendulum of the clock, one of the flowers on the table, and the pair of shoes were deleted.

3. Size change: One item was made larger or smaller. In the bus pictures, the bus was made

50% smaller. In the lady pictures, the vase with flowers was made 50% larger.

4. Rearrangement: The location of one item was changed. In the organized bus picture, the jacket on the hanger was moved from the right to the left of the table; in the unorganized version, the positions of the jacket and the dog were exchanged. In the organized lady picture, the shoes were moved from under the table to the right of the table; in the unorganized version, the positions of the shoes and the car were exchanged.

5. Substitution: One object was replaced by a conceptually similar object. In the bus pictures, the bowl of fruit was replaced by a basket of fruit. In the lady pictures, a different type of clock was used.

Design

The 168 *Ss* were divided into the 3 ethnic groups of 56 *Ss* each. The ethnic groups were further subdivided into equal numbers of boys and girls in the first and second grades. Within each of these 12 groups, each child worked with 2 pictures, an organized picture from one set and an unorganized picture from the other. Order of presentation (organized first or unorganized first; bus first or lady first) was counterbalanced across groups. Presentation order was not a significant variable in any of the analyses and will not be discussed further.

For each of the pictures shown to a child, the following 4 tasks were given in a constant order:

1. Verbal description of the picture while viewing the stimulus.
2. Verbal recall of the picture.
3. *Same–different* recognition test, in which *Ss* were presented with the 5 transformations and the original in a random sequential order. The *Ss* were required to respond to each presentation by saying whether the instance was exactly the *same* as the original or whether it was *different*. If the child responded *different* he was asked to tell how it was different.
4. The paired-comparison recognition test consisted of a 2-alternative forced-choice procedure in which each transformation and the original were paired with all other transformations to make 15 pairs of pictures. The *S* was asked to pick one picture from each pair which looked most like the original picture he had seen and described.

Procedure

Raven's Coloured Progressive Matrices were administered individually by testers of each *S*'s own ethnic group before the experimental procedures began. For this test and the experimental sessions, Chicano *Ss* were given the choice of speaking in English or Spanish. All of them chose to use English, and no language difficulties were observed. Following administration of the Coloured Progressive Matrices, each *S* was tested individually in 2 experimental sessions 1 wk. apart. In each session one target picture was used. At the beginning of the first testing session *S* was told that [illegible] voice would be recorded and that he would be [illegible] to listen to his voice. Then he was told he [illegible] going to see some pictures and that *E* wanted to [illegible] how children looked at pictures and remember [illegible] them. The *Ss* were assured that the testing proce[illegible] had nothing to do with grades or school work [illegible] When *E* felt that some rapport had been gained [illegible] *S* was reasonably relaxed, the experimental p[illegible] dure began.

Description. The *S* was given a practic[illegible] ure which was rich in detail and similar to th[illegible] ermental pictures. The *S* was asked to tell [illegible]y thing he saw in the picture, no matter how [illegible] all the detail, and to leave nothing out. He was [illegible]ed to point to each object or detail as he mentio[illegible] it. If *S* had any difficulty with the describin[illegible] nd pointing procedure, *E* demonstrated what w[illegible] required. The practice picture was kept in fr[illegible] of *S* until he described most of the details i[illegible] he picture.

After this preexperimental procedure, *S* was given the first experimental picture and the de[illegible]ription and pointing instructions were repeated. Presentation time was 3 min. If *S* responded incompletely or stopped too soon, *E* prompted him by asking if he saw anything else. At the end of 3 min., *E* flipped the picture over and talked to *S* for 20 sec.

Recall. The *S* was next asked to recall as many objects and details from the picture as he could remember. When he stopped the first time, *E* prompted him once. If he could not respond, the recall task was ended; otherwise, it continued until he stopped the second time.

Same-Different *task*. Approximately 30 sec after the recall task, *S* was presented with 6 pictures in sequence, 5 of which were transformations of the original picture, and 1 of which was a copy of the original. The *S* was told that he would see 6 pictures, some of which were exactly the same as the picture he had just described and recalled, and some of which were different. If a picture was exactly the same, *S* was to say *same*. If different, *S* was to say *different* and tell exactly how it was different. A different sequential order of pictures was used for each child in a Latin square design.

Paired-comparison task. After a short break of approximately 30 sec., *S* was told he was now going to see pictures 2 at a time. Each time a pair was presented he was to choose the picture which looked most like the original picture he had described and recalled and which was still lying face down on the table. He was told further that if neither picture looked exactly like the original picture he was to choose the one which looked most like the original. The *S* was then presented all combinations of target and transformations, 2 at a time. Presentation sequence was randomly chosen for each *S*.

The second session was identical to the first, except that the practice picture was not presented.

Results

Analysis of variance on the Coloured Progressive Matrices showed that Grade 2

TABLE 1

VERBAL DESCRIPTION AND VERBAL RECALL MEASURES

Variable	Description			Recall				
	No. of items	Inferential + relational statements	No. of repetitions	No. of items	Inferential + relational statements	No. of repetitions	% of description recalled	% of recall not in description
Total (N = 336)	32.4	1.8	4.4	19.4	1.2	2.7	53.5	12.6
Grade								
1	31.1	1.8	4.8	18.1	1.3	2.9	53.4	11.8
2	33.8	1.7	3.9	20.7	1.0	2.5	53.6	13.4
Sex								
Male	34.3	2.1	4.7	21.0	1.4	3.0	53.4	13.6
Female	30.6**	1.5***	4.0	17.8***	.9*	2.5	53.5	11.5
Ethnic								
White	36.1	2.2	4.4	21.1	1.4	2.9	53.1	12.2
Black	32.1***	2.1	6.1***	19.4*	1.4	3.2*	53.0	13.0
Chicano	29.0***	1.1***	2.6***	17.7*	.7*	2.1*	54.3	12.5
Organization								
Organized	32.0	1.8	4.4	20.8	1.4	3.0	57.0	13.7
Unorganized	32.9	1.8	4.4	18.0**	1.0	2.4	49.9***	11.5*
Set								
Lady	30.8	2.2	4.5	19.1	1.5	3.0	55.4	11.3
Bus	34.0*	1.4***	4.2	19.8	.8***	2.5	51.6*	13.8*

* Group values differ significantly, $p < .05$.
** Group values differ significantly, $p < .01$.
*** Group values differ significantly, $p < .001$.

scored higher than Grade 1, F (1, 156) = 19.25, $p < .001$, and males scored higher than females, F (1, 156) = 22.60, $p < .001$. There were no main effects due to ethnic group, but there was a Grade Level × Ethnic Group interaction, F (2, 156) = 8.27, $p < .01$. Newman–Keuls tests showed that black children in the first grade scored lower than the other groups, which did not differ from each other.

Verbal Description and Recall

Verbal description and recall measures are summarized in Table 1. The first column shows the main description score, consisting of the total number of items mentioned (either whole objects or details). Inferential statements (e.g., the girl *is happy*) and relational statements (e.g., the bowl *is on* the table) as well as repetitions were scored separately and not counted in the main description score. These measures are shown in Columns 2 and 3, respectively. Similar analyses were carried out for verbal recall, shown in Columns 4–6. Two other recall measures were computed: the percentage of items in each *S*'s description that also appeared in his recall, and the percentage of items in recall that had not been mentioned in that *S*'s description. These measures are shown in the last 2 columns. Analyses of variance (3 Subject Variables × 2 Stimulus Variables) were carried out on the various verbal measures. Significance levels for the analyses are also shown in Table 1.

Subject variables. (*a*) Grade—there were no main effects on any verbal measure due to grade level. (*b*) Sex—males had higher description and recall scores than females. Interactions with grade level on number of items in description ($p < .05$) and in recall ($p < .01$) showed that this effect was due primarily to the second-grade boys, who described and recalled significantly more than the other groups. (*c*) Ethnic Group—whites both described and recalled more than blacks, who in turn described and recalled more than Chicanos.

Stimulus variables. (*a*) Organization—degree of organization of the pictures did not affect amount of description but did affect amount of recall. Organized pictures produced a greater number of items recalled than unorganized pictures. (*b*) Picture set: The bus pictures produced more items in description than the lady pictures, but there was no significant difference in amount of recall of the 2 kinds of picture.

Number of inferential and relational statements was small in comparison with number of items mentioned, perhaps because of the type of instructions given, and the pointing method of description used. The 2 types of statements were combined into a single measure for purposes of analysis. In general, scores on this measure paralleled the total description and recall scores, males making more inferential and relational comments than females, and Chicanos making fewer than blacks and whites, who did not differ from each other.

Repetition of items occurred fairly frequently in both description and recall. There were no instructions that *S*s should not repeat themselves, but these data may be contrasted with the lack of repetitions, also without instructions, found in adult recall of verbal material (Borges, 1972). There were no significant differences in amount of repetition in description and recall as a function of stimulus variable, and the only *S* variable affecting repetition was ethnic group. Blacks repeated items more than whites, and Chicanos had very low scores on this measure.

There are several ways in which recall might be expressed in relation to description. Column 7 of Table 1 shows the percentage of items described that were also recalled. The mean for the total sample on this measure was 54%, a score high in comparison with immediate recall of well-organized lists of words (Mandler, Pearlstone, & Koopmans, 1969). However, recall can also be represented as a percentage of total number of items described by the *S* pool. It can be argued that the total number of items in a picture is better represented by the "total possible" set of items, rather than the subset actually mentioned by *S*. Such an assumption is bolstered by the finding that an average of 13% of an *S*'s recall score consisted of items not mentioned in his description of the picture. An inventory of all objects and details mentioned by any *S* (excluding different labels for the same items) was constructed for each picture, resulting in a total of 129 items in the lady pictures and 152 items in the bus pictures. Recall, expressed as a percent of this inventory, drops to a mean of 14% correct. Neither of the above procedures seems an entirely satisfactory method of estimating amount of recall; at best the 2 estimates provide upper and lower limits on amount of verbal recall of complex pictures.

Finally, it should be noted that there was a high *S* correlation between number of items mentioned in description and in recall, $r = .73$, $p < .01$. Both verbal measures had a low correlation with Coloured Progressive Matrices, $r = .10$, *ns*, and .16, $p < .05$, respectively.

Same–Different *Recognition Test*

Table 2 shows the proportion of correct recognition of the original picture, each of the transformations, and the total proportion correct. A rough correction for possible response bias in saying same or different was calculated by multiplying overall rate of saying *same* or *different* by actual occurrence of same or different pictures. This method resulted in an estimate of chance probability of a hit on the transformations of .48, and chance probability of a hit on the original of .07. It can be seen from Table 2 that recognition of the substitution transformation was essentially at chance, with size and rearrangement resulting in somewhat higher scores. Recognition was best for deletion, reversal, and the original picture. Overall, probability of correct recognition was .70.

Analyses of variance (3 Subject Variables × 2 Stimulus Variables) were carried out on a total recognition score, consisting of the total number of correct recognitions, and separately on the number correct on each transformation. Significance levels for differences among the various groups are also shown in Table 2.

Subject variables. (*a*) Grade—Grade 2 had a significantly higher total recognition score than Grade 1. This difference was primarily due to Grade 2 having higher scores on the reversal and deletion transformations. Thus, Grade 2 showed superior recognition even though there were no differences between the grades on verbal description or recall. (*b*) Sex—there were no significant differences between total

TABLE 2

Proportion Correct on Original Pictures and Five Transformations in *Same–Different* Recognition Test

Transformations	Original	Substitution	Size	Rearrangement	Deletion	Reversal	Total
Total (N = 336)	.89	.49	.58	.59	.77	.91	.70
Grade							
1	.92	.45	.54	.55	.68	.87	.67
2	.86	.53	.61	.62	.86**	.95*	.74*
Sex							
Male	.87	.50	.58	.57	.77	.92	.70
Female	.91	.48	.57	.61	.77	.90	.71
Ethnic							
White	.88	.52	.58	.62	.76	.96	.72
Black	.87	.50	.59	.59	.77	.86**	.69
Chicano	.92	.45	.56	.54	.79	.91**	.70
Organization							
Organized	.91	.51	.62	.64	.71	.93	.72
Unorganized	.87	.47	.53**	.54**	.84**	.89	.69***
Set							
Lady	.91	.61	.52	.73	.80	.93	.75
Bus	.87*	.37***	.64	.45***	.74*	.88	.66***

* Group values differ significantly, $p < .05$.
** Group values differ significantly, $p < .01$.
*** Group values differ significantly, $p < .001$.

correct or on any transformation between males and females. Again, this result may be constrasted with the results on the verbal measures, in which males showed higher description and recall scores. (*c*) Ethnic group—there were no significant differences in total number correct. This result may also be contrasted with the differences found among ethnic groups on description and recall. The only significant difference in recognition occurred on the reversal transformation, on which blacks were slightly less accurate. A Grade × Ethnic Group interaction on this transformation ($p < .01$) indicated that the difference in recognition occured only in the first grade. Lower scores on the Progressive Matrices were also found for the first-grade black sample. Although this test is usually used as a nonverbal IQ test, it has a strong spatial component.

For all 3 S variables, verbal description and recall measures were poor predictors of accuracy of recognition. An S correlation between number of items described and total correct recognition resulted in $r = .21$, $p < .01$, and between number of items recalled and total correct recognition, $r = .25$, $p < .01$. The correlation between Coloured Progressive Matrices and total correct recognition was .16, $p < .05$.

Stimulus variables. (*a*) Organization—total correct recognition was significantly higher for organized pictures. The main transformations involved in this difference were rearrangement and size. These 2 transformations are most directly relevant to the degree of organization of a picture as it has been defined here, namely, as the number of meaningful relationships among the items in the picture. Both the rearrangement and size transformations involve a change in relationships among objects, and these changes were better recognized in the more organized pictures.

The other major difference occurred on the deletion transformation. Deletion of detail was better recognized in the unorganized pictures. Since there were no meaningful relationships among the objects in the unorganized pictures, it is possible that more attention was paid to details of the objects represented. In the organized pictures, on the other hand, more processing of relationships among objects took place. This does not appear to be a question of how Ss spent a limited processing time, since presentation of each picture lasted 3 min., but rather a question of what aspects of the organized and unorganized pictures tended to be coded for storage. When a deletion transformation was recog-

nized as different, on the average only 1 of the 3 deleted items was mentioned; in almost all cases this was the book in the bus pictures and the shoes in the lady pictures. In both cases these items were the largest of the 3 deletions. Apparently the 2 small deletions of detail were rarely noticed, although more often in the unorganized pictures.

(*b*) Picture set—The lady pictures were recognized significantly more often than the bus pictures. It will be recalled that the lady pictures contained less information, in the sense that they produced less verbal description and had a smaller total inventory of items. Although it has been assumed that high information pictures are easy to recognize because distractors tend to be very different, when distractors consist of small changes on the target pictures it is possible that recognition is hindered by a greater amount of information. Amount of information would have to be systematically varied to test this hypothesis.

There were significant interactions between the set and organization variables on the substitution, rearrangement, and size transformations. These differences appeared to reflect uncontrolled differences in saliency of various objects in the different pictures, perhaps in part as a function of their location in the picture as well as degree of organization in the picture as a whole. Since parametric variation in transformations of objects was not carried out in this experiment, no assessment of relative importance of, for example, size changes vs. substitutions, in recognition of meaningful pictures in general can be made at this time. It is likely, in fact, that accuracy of recognition of any given transformation can be systematically varied by manipulating the saliency of the object which is transformed and the extent of the transformation (e.g., amount of size change, number of rearrangements, etc.).

Concerning the relationship between verbal recall and recognition as a function of stimulus variables, only degree of organization affected both measures. More was recalled about the organized pictures, and they were better recognized. These differences in memory were not, however, related to number of words used to describe these types of picture.

Paired-Comparison Test

The *same-different* test showed which transformations were most difficult to recognize, i.e., which were more similar to or confusable with the original picture. The question then becomes whether it is possible to scale distances along a dissimilarity dimension(s) to aid our understanding of the structure of what is stored in memory. If it is assumed that memory for pictures can be represented as a multidimensional Euclidean space, and that the judged degree of similarity of any transformation to the original or to other transformations is inversely related to distances within that space, then the paired-comparison test can be used to estimate where in that space each transformation of the original picture is located.[4]

The mean percent times that the original and each transformation was chosen in comparison with every other was computed separately for the various groups and converted to d' scores (a measure of discriminability). Thus, matrices of d' scores were formed from the 15 paired comparisons of each original picture and its 5 transformations. These matrices give an estimate of the distance each transformation is from each other and from the original in the hypothetical memory space. To the extent that the original can be considered as an anchor point in the space (since each paired comparison was made on the basis of similarity to the original), the distances from the original to the transformations can also be considered a measure of distortion of the original by the particular transformation made.

Independent matrices, each based on 168 observations, were constructed for the 2 sets of original pictures, for the organized and unorganized versions of the pictures,

[4] An example of a similar type of analysis used to explore the structure of semantic, rather than pictorial, space can be found in Henley (1969), and for discriminability of letterlike forms in Harrington and Brown (1972).

and for Grades 1 and 2. Matrices for the 3 ethnic groups were based on 112 observations each. The breakdown by sex was not included because of the lack of differences found on the *same–different* test. Each of the matrices was analyzed separately by the TORSCA multidimensional scaling program (Young & Torgerson, 1967).

The Kruskal stress index for each analysis (giving an estimate of goodness of fit of the various points in a given number of dimensions) for 1- and 2-dimensional solutions, as well as percentage of variance accounted for by the dimensions are shown in Table 3. In all cases except the Chicano group, the 2-dimensional solution gave close to perfect fit (stress varying .01%–.6%), and even in the Chicano group the fit was excellent (1.6%). For the 1-dimensional solutions, stress varied .01%–10.4%. Since only 6 items were being compared, the excellent fit in 2 dimensions might have occurred by chance, although the high degree of similarity among solutions obtained from independent groups of *S*s and between different pictures makes such a possibility unlikely. Based on Klahr's (1969) analyses, the probability of obtaining by chance stress indices as low as those in the 1-dimensional solutions is less than .05 in each case.

The 1-dimensional solutions were highly similar to the first dimension of the 2-dimensional solutions, presumably because in both cases one dimension accounted for most of the variance. Although the 1-dimensional solutions were adequate descriptions of the data, the 2-dimensional solutions are presented here because they give a clearer picture of the data, and because work in progress with adults suggests that 2-dimensional solutions will be necessary. Figure 2 shows the solutions obtained from the various subgroups. For each breakdown the groups being compared are shown in the same space, although they were derived independently. Inspection of Figure 2 shows that in all cases original is located at one end of the main dimension, and that reversal is located at the greatest distance from the original, with deletion the next farthest from the original. The other 3 transformations are located in between, show the most overlap, and the most movement from analysis to analysis. No attempt was made to conceptualize the second dimension since it accounted for relatively little of the variance, although in all cases deletion was at one end of the dimension and rearrangement at the other.

TABLE 3

PERCENT STRESS AND PERCENTAGE OF VARIANCE ACCOUNTED FOR BY EACH DIMENSION IN ONE- AND TWO-DIMENSIONAL SOLUTIONS

Variable	Solution: Two-dimensional: Stress	Two-dimensional: Variance: First dimension	Two-dimensional: Variance: Second dimension	One-dimensional: Stress	One-dimensional: Variance
Set					
Lady	.01	79.8	20.2	.01	100.0
Bus	.12	90.2	9.4	.01	100.0
Organization					
Organized	.56	89.3	9.6	10.40	87.0
Unorganized	.05	89.1	10.9	5.07	91.4
Grade					
1	.01	93.8	6.2	.01	100.0
2	.01	85.8	14.2	9.52	83.1
Ethnic					
White	.01	90.4	9.6	.01	100.0
Black	.03	93.6	6.4	6.10	90.2
Chicano	1.61	89.5	8.2	6.67	92.2

Panel 1 shows the comparison between the 2 types of original picture. There should be a close correspondence between these 2 TORSCA analyses if we wish to conclude that the pictorial memory space thus constructed is general and not idiosyncratic to a particular picture. Looking at Panel 1 we can see that original, reversal, and deletion occupy roughly the same spaces. The greatest change from one picture to the other occurs on size and rearrangement, followed by substitution. Comparing the movement of these 3 transformations (toward or away from the original as a different picture is used) with the proportion of correct recognition for these transformations in the *same–different* test, shown in Table 2, we can see that in each case in which a transformation was

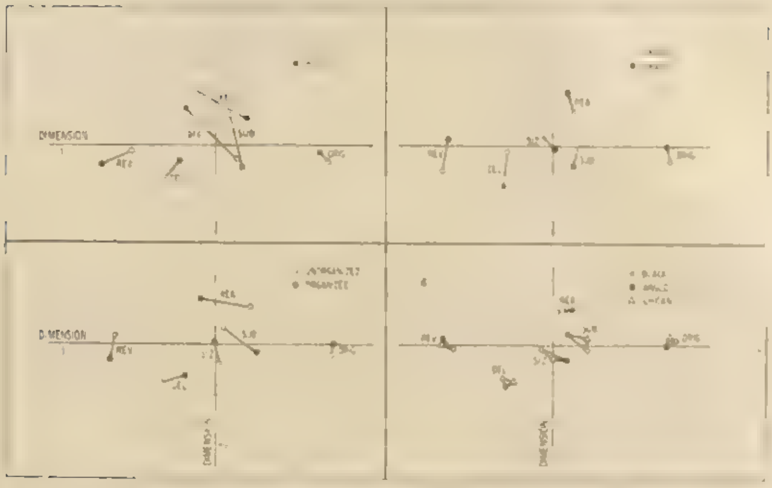


FIGURE 2. Two-dimensional TORSCA multidimensional scaling analyses derived from independent groups of Ss and different types of stimuli. (Panel 1 compares the 2 types of picture, Panel 2 compares organized vs. unorganized versions of the pictures, Panel 3 compares 2 grade levels, and Panel 4 compares 3 ethnic groups. Stress indices are given in Table 3. Abbreviations: REA = rearrangement, SIZ = size, SUB = substitution, REV = reversal, DEL = deletion, and ORIG = original.)

recognized less well it is closer to the original in the TORSCA analyses. As mentioned above, these 3 transformations seem to have different effects on recognition depending upon the saliency and location of the particular item transformed in a given picture.

Turning to the effects of organization, Panel 2 indicates that the rearrangement transformation was most affected by differences in degree of organization. Rearrangement was less frequently recognized in the unorganized pictures in the *same–different* test, and in the paired-comparison test this transformation was judged to be closer to the original in the unorganized pictures. Thus, there is good agreement between the 2 tests on the greater importance of relationships among objects in organized pictures.

Comparisons between Grades 1 and 2 are shown in Panel 3. They are very similar to each other on the first dimension. The first graders in fact show essentially a 1-dimensional solution, while the second graders' scores have begun to move into the second dimension. Although current work with adults indicates that the second dimension accounts for a still greater part of the variance, more work [illegible] led to conclude that there is a d[illegible]ntal difference in complexity of pictori[illegible]emory space.

Panel 4 shows comparisons among the 3 ethnic groups. The chief conclusion is that the pictorial memory space is highly similar for the 3 groups. This result confirms the lack of difference among ethnic groups on the *same–different* test.

Overall, there was excellent agreement between the 2 types of recognition test. The advantage of the paired-comparison test is that it gives a more detailed picture of differences between the original and the transformations. In addition it eliminates response biases such as the tendency to say *same* or *different* to most pictures, a factor which might well vary among Ss of different ages. Although a left- or right-choice bias could exist in the paired-comparison test, analyses of the number of times Ss chose the left or right picture indicated that no such bias was operating.

Because the 2 recognition tests were given in a constant order, it is possible that performance on the *same–different* test influenced choices on the paired-comparison test. For example, if an S correctly recog-

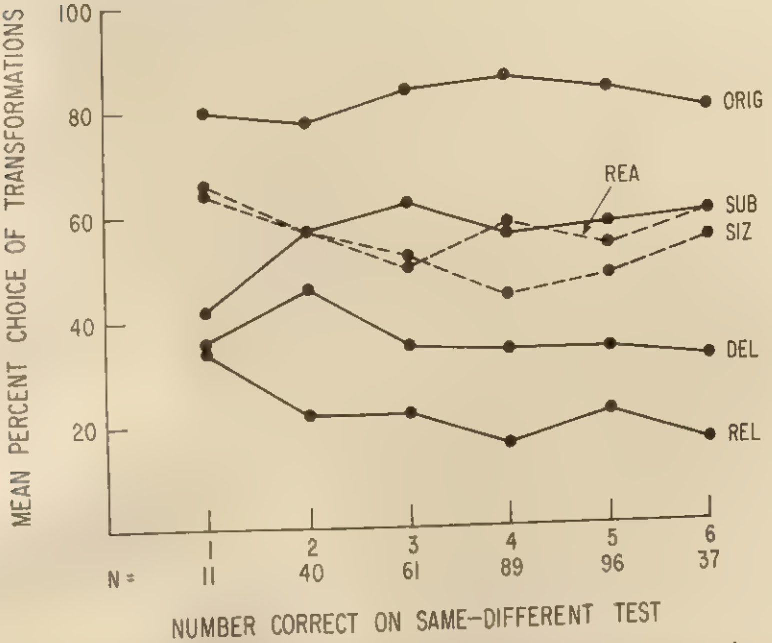


FIGURE 3. Percent choice of each transformation in the paired-comparison test plotted as a function of number of correct responses on the *same–different* test. (Number of *Ss* involved in each point is shown below the abscissa.)

nized a transformation as different, he might be less likely to choose it in the paired-comparison test. To check on this possibility, percent choice of each transformation in the paired-comparison test was plotted as a function of the number of correct responses on the *same–different* test. Two *Ss* who made 6 out of 6 errors on the *same–different* test were excluded from this analysis. The remaining *Ss*, grouped according to number correct, are shown in Figure 3. It can be seen that performance on the paired-comparison test did not differ as a function of accuracy on the *same–different* test, indicating at least some degree of independence of the 2 tasks.

Discussion

The present data offer little support for a conception of unlimited pictorial recognition. In spite of long presentation times, overall recognition rate was fairly low (70%) and varied from chance responding to 91% correct depending upon the nature of the distractor. The distractors used in this experiment consisted of transformations on the original pictures and thus were similar to the target pictures. Conflicting results have been reported when similar distractors have been used, some studies finding lower recognition rates, some not. The question then becomes how to define and scale similarity of complex pictures which contain many visual and semantic dimensions.

The multidimensional scaling method used in this experiment appears to be a useful technique for scaling the similarity of distractors to the picture stored in memory, and in addition to discover which aspects of pictures are most likely to be retained. Not only were transformations on the target pictures stably located in a memory space, but changes within that space were related to type of picture and the degree of organization within pictures. For example, the judged similarity of a given distractor differed as a function of amount of organization in the stimulus materials.

There was a very close correspondence between the 2 types of recognition tests, even though they asked for different judgments. In the *same–different* test, *S* was asked whether he had seen a picture before, and in the paired-comparison test he was required to make a similarity judgment. Although it has been assumed here that similarity of distractor to

target is a major factor in recognition accuracy, it should be pointed out that this assumption needs further testing. In the first place, similarity judgments may differ when they are made in the presence of the target items than when they are made about remembered targets. Second, it could be the case that reversal of a picture, for example, is easy to recognize as different yet be judged most similar to the original on the grounds that all the relationships among the objects are the same. The fact that recognition and similarity judgments were highly correlated for children does not necessarily mean that the same relationships will hold for adults. This problem is currently being investigated.

Although the present data do not allow an assessment of the relative importance of various transformations in recognition of pictures in general, the results for 2 of the transformations should be mentioned. Two transformations were consistently recognized across pictures and also judged as most dissimilar, namely, reversal and deletion of detail. Although experiments with adult *S*s have typically found only slight loss in recognition accuracy when reversals were used (Dallett et al., 1968; Standing et al., 1970), the .91 rate of recognition of reversals in this experiment seems surprising in light of a number of studies showing that children have difficulty in discriminating among reversals of pictorial geometric forms (e.g., Rudel & Teuber, 1963) Ease in recognizing reversals in more complex pictures may be a function of the total number of items whose location is changed. In the present experiment, reversal involved a greater number of locational changes than any other transformation. On the other hand, recognition of reversals may be a function of the meaningfulness of the material used. Current work is being done to evaluate these possibilities.

The other transformation most consistently recognized as different was deletion of detail. This result is consistent with experiments reported by Vurpillot (1972), showing that children of age 7 are better at detecting differences in detail in meaningful pictures than size changes or rearrangement. On the other hand, her experiments indicated that children of this age are also good at detecting small substitutions, a result at variance with the present data. It should be pointed out, however, that it is difficult to compare results on transformations from different studies because of the lack of appropriate scaling of the magnitude of the changes involved. In addition, Vurpillot's work, like that of Gibson, Gibson, Pick, and Osser (1962), involved detection of differences in a matching-to-sample task. Little is known about attentional or search patterns in detection vs. memory tasks, or whether the differences easiest to detect are those which are most often retained in memory.

Concerning the relationship between verbal measures and recognition in the present experiment, for the *S* variables, verbal measures were poor predictors of accuracy of recognition. This finding is of particular interest in light of the heavy verbal loading found in the test batteries most frequently used to assess perceptual functioning in minority children. In the present study, black and Chicano children said less than white children, but the volume of words bore no relation to the visual recognition measures. In terms of accuracy of recognition, responsiveness to the organization of the pictorial materials, and types of mistakes made on transformations, the 3 ethnic groups were highly similar. The multidimensional scaling of the similarity judgments produced a pattern of responding that was virtually identical for the 3 ethnic groups

The stimulus variables showed 2 types of relationship between verbal and visual measures. The pictures that contained more information (had a higher inventory of items) called forth more words in description and were also more difficult to recognize. Thus, there was a negative relationship between amount contained in a picture and amount remembered. On the other hand, the organized and unorganized versions of the pictures, which contained equal numbers of items and called forth the same number of words in description, nevertheless showed differences in both recall and recognition as a function of degree of organization. Thus, amount of organizational or structural information showed a positive relationship with amount remembered. It should be noted that an hypothesis that equates amount of organization with number of words used in description, such as the verbal-loop hypothesis (Glanzer & Clark, 1964), is not a useful equation in the type of pictorial recognition studied in this experiment.

In summary, the data of this study indicate that children are responsive to the organization of pictorial materials and that the nature of the stored representations reflects the amount of organization in the material viewed. There was little verbal indication of such responsiveness; that is, very few comments

were made in description or recall about relationships among the objects in the pictures, nor were more or different words used to describe organized scenes than unorganized scenes. However, information about relationships among the objects in the organized scenes was processed and retained for storage. When unorganized pictures were presented, which did not include meaningful relationships among the objects, other types of information, such as characteristics of small items, were processed and retained. Although we are still far from being able to specify the nature of the stored representations of pictorial material, the results of the multidimensional scaling technique suggest that useful models of memory representations can be constructed that will contribute to the development of a theory of pictorial memory.

REFERENCES

Bahrick, H. P., & Boucher, B. Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 1968, **78**, 417–422.

Bahrick, H. P., Clark, S., & Bahrick, P. Generalization gradients as indicants of learning and retention of a recognition task. *Journal of Experimental Psychology*, 1967, **75**, 464–471.

Borges, M. A. Increasing item accessibility in free recall. *Journal of Experimental Psychology*, 1972, **95**, 66–71.

Brown, A. L., & Campione, J. C. Recognition memory for perceptually similar pictures in preschool children. *Journal of Experimental Psychology*, 1972, **95**, 55–62.

Brown, A. L., & Scott, M. S. Recognition memory for pictures in preschool children. *Journal of Experimental Child Psychology*, 1971, **11**, 401–412.

Cohen, J. The factorial structure of the WISC at ages 7–6, 10–6, and 13–6. *Journal of Consulting Psychology*, 1959, **23**, 285–299.

Dallett, K., Wilcox, S. G., & D'Andrea, L. Picture memory experiments. *Journal of Experimental Psychology*, 1968, **76**, 312–320.

Daniel, T. C. Nature of the effect of verbal labels on recognition memory for form. *Journal of Experimental Psychology*, 1972, **96**, 152–157.

Farnham-Diggory, S. Cognitive synthesis in Negro and white children. *Monographs of the Society for Research in Child Development*, 1970, **35**(Serial No. 135).

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Gibson, E. J., Gibson, J. J., Pick, A. D., & Osser, H. A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 897–906.

Glanzer, M., & Clark, W. H. The verbal-loop hypothesis: Conventional figures. *American Journal of Psychology*, 1964, **77**, 621–626.

Goldstein, A. G., & Chance, J. E. Visual recognition memory for complex configurations. *Perception & Psychophysics*, 1971, **9**, 237–241.

Haber, R. N. How we remember what we see. *Scientific American*, 1970, **222**, 104–112.

Harrington, L. F., & Brown, D. R. Visual pattern perception: Multidimensional analysis of discrimination difficulty for children. *Psychonomic Science*, 1972, **29**, 219–221.

Henley, N. M. A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 176–184.

Klahr, D. A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. *Psychometrika*, 1969, **34**, 319–330.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 410–423.

Nelson, K. E. Memory development in children: Evidence from nonverbal tasks. *Psychonomic Science*, 1971, **25**, 346–348.

Pettigrew, T. F. *A profile of the Negro American.* Princeton, N. J.: Van Nostrand, 1964.

Rosinski, R. R. The development of attention and recognition. Unpublished doctoral dissertation, Cornell University, 1970.

Rudel, R. G., & Teuber, H. L. Discrimination of direction of line in children. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 892–898.

Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

Standing, L., Conezio, J., & Haber, R. N. Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 1970, **19**, 73–74.

Tyler, L. E. *The psychology of human differences.* (2nd ed.) New York: Appleton-Century-Crofts, 1956.

Vurpillot, E. *Le mode visuel du jeune enfant.* Paris: Presses Universitaires de France, 1972.

Young, F. W., & Torgerson, W. S. TORSCA, A FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. *Behavioral Science*, 1967, **12**, 498.

(Received July 17, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 670–677

INSTRUCTIONAL CONTROL OF SERIAL-LEARNING STRATEGIES[1]

WESLEY A. KAYSON[2]
Iona College

WILMA A. WINNICK
Queens College, City University of New York

Two sets of instructions were used in a serial-learning task to induce one group to form sequential associations and a second to use the position strategy. Performance was contrasted with control groups, and half of the *S*s in each of the 3 groups learned lists of high-association trigrams, with the other half learning low-association lists. Examination of the latencies with which responded, after learning, with the trigram that followed the presented or with the trigram at the serial position presented provided evidence that instructed groups had used the suggested strategies, but gave no indication of the consistent use of either strategy by control *S*s. Analysis of the error data found a significant effect of instruction, with positional instruction superior, and a significant interaction between instructions and association value, with position instructions highly effective for the low-association list

Recent attempts to delineate the effective stimulus underlying serial verbal learning have focused upon 3 possible cues for the anticipatory response required in learning of this type: (*a*) the immediately preceding item, (*b*) a combination of 2 or more preceding items (an elaboration of the first), or (*c*) the identification of the ordinal position of the item. Young (1962) has viewed these as possible hypotheses that *S* might adopt in serial learning (SL).

Most attempts to identify the stimulus in serial learning have proceeded by utilizing transfer techniques to seek out indications of the cues used in serial learning. In such studies, learning of a serial list is typically followed by a second learning task, the aim of which is to detect the indicators of the cues used in SL. Thus, the amount of transfer has been assessed from the learning of one serial list to the learning of a second list in which the items from the original list occupy the same or different positions; this technique was used by Young (1962) to evaluate the positional hypothesis. More commonly used is transfer from SL to paired-associate learning (PAL) where, for example, items that were contiguous in previous SL are used to form pairs for PAL: learning of such items is compared with the learning of items that had not been in sequence in SL (Young, 1962). In another variation of the transfer technique, SL is the second task, and the first learning encourages or suggests cues (usually positional) that might aid or hinder subsequent SL. In a test of the efficacy of the positional cue, Ebenholtz (1963) studied transfer from PAL, where spatial position cues were paired with verbal items, to SL, in which items occupied the same or different spatial positions. In another PAL–SL transfer study, Winnick and Dornbush (1968) followed PAL, in which syllables were either in the positions indicated by the numerals or in other positions. These experiments and others summarized by Young (1968) lead one to agree with Young that the issue of effective stimulus for serial learning has not yet been resolved.

The present study proposed to view the cues mentioned as providing the basis for a plan or strategy of learning that might be consciously chosen by *S*, suggested by the experimental materials, or dictated by the instructions employed in the experiment. In contrast with previous studies—most of which have searched, after learning has taken place, for evidence indicating the use of one or the other of these cues—the present study has attempted, by the instructions used, to induce *S*s to employ one or the other cue as a strategy for the learning of the serial list.

---

[1] This article is based upon Experiment I of a dissertation submitted by the first author in June 1970 to the City University of New York in partial fulfillment of the requirements for the PhD degree.

[2] Requests for reprints should be sent to Wesley A. Kayson, Department of Psychology, Iona College, New Rochelle, New York 10801.

In an attempt to determine the relative effectiveness of such strategies, the present study has used instructions to suggest the strategy of tying successive items together (Young's 1962 serial hypothesis) or of identifying the ordinal position of the items (the positional hypothesis). The possibility that the effectiveness of a particular approach might be functionally related to the particular materials learned was examined by the use, in separate groups, of lists composed of either low-association (LA) or high-association (HA) nonsense syllables. It seemed reasonable to expect that HA syllables would provide associative bases for linkages, and that the serial strategy would therefore prove more feasible for these materials. By contrast, the lack of potential in LA nonsense syllables for associative linkages suggested that the positional strategy might be more effective for these materials.

Method

Subjects. Ninety volunteer *S*s from introductory psychology courses in the Queens College summer session participated in the experiment. They were individually assigned to 1 of the 6 groups in the order of their appearance in the laboratory, such that the number of *S*s per group was equal after every sixth *S*. Two *S*s were dropped from the experiment, one for failing to follow instructions and one due to *E* error. One of the dropped *S*s was in the HA-position group and the other in the LA position group.

Apparatus and materials. The stimulus items consisted of 2 sets of consonant-vowel-consonant (CVC) trigrams selected from Archer's (1960) list. The LA syllables were between 9% and 41% association value, and the HA syllables were between 73% and 95%. To control for any possible effects due to item ordering within the lists, 3 different orders of these syllables were formed, so that pairs of such HA and LA lists provided 3 replications of the experiment. The syllables were typed in capital letters on a standard memory-drum tape and presented on a Lafayette memory drum at a 2-sec. rate. The same interval was used between movements of the memory drum to pace recall. There was an intertrial interval of approximately 30 sec.

Experimental design. Six independent groups of *S*s were individually tested. One group was instructed to learn the list by associating each nonsense syllable with the following one (Group S). The second was told to learn the list by associating each nonsense syllable with the number of its position in the list (Group P). The third (control) group was given only standard instructions, with no suggestions about how to learn the list (Group C). Half of the *S*s in each group were presented with a list composed of HA syllables, and half learned lists composed of LA syllables. Randomly selected pairs of HA and LA lists formed 3 replications of the basic experiment. Thus, a $3 \times 3 \times 2$ factorial design was employed, with 3 levels of instruction, 3 replications, and 2 levels of associative value. The 6 distinct groups were designated as S–HA, S–LA, P–HA, P–LA, C–HA, and C–LA.

Procedure. All *S*s served individually, learning the list to a criterion of one perfect trial by the serial recall method for "the identification and assessment of processes in serial learning [Battig & Lawrence, 1967, p. 178]." In this method, learning trials which exposed the items alternated with blank exposures during which *S*s were to attempt recall. Before the first learning trial, each *S* was given the following modified standard learning instructions:

> This is a memory test. You are to learn the nonsense syllables that appear in the window of the machine in the right order as quickly as you can. When I start the machine the syllables will appear in the window one after another until 12 have been shown. Try to learn them as they appear.

After the twelfth syllable was shown and before the first recall trial, the following instructions were given:

> Now, right after an asterisk has appeared in the window a blank line will appear. Then you are to say what the first syllable is. Then another blank line will appear in the window. While that line is there, you are to say what the second syllable is. You are to continue in this manner until 12 lines have appeared in the window. Guess even if you are not sure.

After the first recall trial, *S* was told:

> Now the 12 syllables will be shown to you again for you to learn. This will again be followed by 12 blank lines. The procedure will continue in this way until you have learned all 12 syllables.

In addition, Group S was told:

> It has been shown that people learn this list faster if they form associations between the syllables. In other words they associate the first syllable with the second, the second with the third, and so on. For example, if the first syllable is "zop" and the second "kip," try to make an association between "zop" and "kip." If the third syllable is "buf," also try to make an association between "kip" and "buf." Remember, try to form associations between the syllables.

Group P was told:

> It has been shown that people learn this list faster if they remember the number of the syllable on the list. In other words, if "zop" is the first syllable, "kip" the second and "buf" the third,

say to yourself "first is 'zop,' second is 'kip,' third is 'buf,' " and learn the entire list in this manner. Remember, learn the entire list by remembering the number of the position that the syllable occupies on the list.

The *S*s in Group C were given only the revision of the standard serial instructions. In addition, every few trials, *S*s in Groups S and P were reminded during the intertrial interval of the method that they were to use to learn the list. The CVC syllables used in the instructions were all different from the syllables used during the learning task.

After the list was learned, each experimental *S* was given 2 kinds of reaction time tests to determine whether he had followed the instructions while learning the list. Each control *S* was also given both reaction time tests in an attempt to infer the strategy used. The equipment consisted of a standard reaction time key attached to a Hunter timer. The reaction time for each trigram was measured to the nearest hundredth of a second. When *E* read the appropriate instructions, he depressed the key and held it down until *S* gave a response, at which time he removed his finger and recorded the reaction time. The 2 sets of instructions that were given to each *S* were as follows: (*a*) "Now I am going to test you on how well you have learned the list. Would you please tell me which syllable is number ______?" These instructions were repeated for each serial position, with 3 different orders of randomly chosen numbers. (*b*) "Now I am going to test you on how well you have learned the list. Would you please tell me which syllable comes after ______?" These instructions were repeated for each trigram. Three different orders of randomly chosen trigrams were used as stimuli.

Half of the *S*s in each group received Instruction *a* first until 12 numbers were presented and then Instruction *b* until the 12 trigrams were presented. The order of instruction presentation was reversed for the remaining *S*s. All *S*s were given both sets of reaction time instructions.

It would be expected that *S*s given P instructions would make a stronger association between the number presented and the appropriate trigram than between successive trigrams. This stronger association should be manifested by a faster reaction time for Reaction Time Instruction *a* than for Reaction Time Instruction *b*, and Group P should therefore respond to Instruction *a* more rapidly than to Instruction *b*. Conversely, it would be expected that Group S would make stronger associations between successive trigrams than between the number of the trigram on the list and the trigram. This stronger association should result in more rapid responding to Reaction Time Instruction *b* than to Reaction Time Instruction *a*. The same reasoning should help infer which strategy control *S*s had used. The above should hold true for trigrams that are recalled correctly. When *S* answers incorrectly to either set of reaction time instructions, there is no way of knowing how strong the association is. Therefore, only trigrams which were responded to correctly after both Instructions *a* and *b* [illegible] used in this analysis.

Results

Reaction times. The first step in the analysis of data was a comparison of the reaction times obtained from the 6 groups. It was reasoned that if the instructions were followed, strong associations would be built up between successive trigrams for the S groups and between the number denoting position of the trigram on the list and the trigram for the P groups. Such strong associations should result in faster reaction times after learning when the "appropriate" rather than the "inappropriate" reaction time instructions were given.

Table 1 shows the mean reaction times for the 6 groups for the preceding stimulus items and the position numerals. The data are broken down into types of reaction time and the order of recall for HA and LA trigrams. Of the 30 HA experimental *S*s, 23 had faster reaction times when reaction time instructions were appropriate for the experimental instructions than when they were inappropriate. The probability of this occurring by chance according to the sign test is less than .004. The apparently slower reaction time for *S*s given the position (appropriate) reaction time instruction first ($\bar{X}$ = 4.13 sec.) as compared to the serial (inappropriate) instructions ($\bar{X}$ = 3.17 sec.) is due to the extreme delay in responding of one *S* in this group who took an average of 17.21 sec. to respond to serial reaction time instructions. In spite of this *S*, 7 out of 9 *S*s responded more rapidly to position reaction time instructions than to serial reaction time instructions. When reaction time instructions were appropriate for the experimental instructions, 22 of 30 LA experimental *S*s had faster reaction times than when the reaction time instructions were inappropriate. The sign test indicated that the probability of this occurrence was less than .008. It would appear that experimental instructions were followed by most *S*s.

It was also argued that it would be possible to infer the strategy used by control *S*s by comparing their reaction times

for position and serial recall instructions. Ten of the 15 LA control *S*s had faster reaction times when position reaction time instructions were used. The probability of this occurrence is equal to .15. Eight of the 15 HA control *S*s had faster reaction times when serial reaction time instructions were used. The probability of this occurrence is .50. Therefore, no assumptions regarding the preference of strategies for control *S*s are warranted.

Finally, it can be seen that the second time *S*s were asked to recall the syllables, their reaction times were faster than the first time they were asked to recall the syllables. It seems that *S*s did not expect to be tested on their recall of the syllables, and it took some time before they became adapted to the reaction time task. The finding that it takes *S* some time to adapt a set to respond rapidly is quite common in reaction time studies.

Since these data indicate that *S*s followed instructions, it is appropriate to evaluate the effect of differential instructions and association values by analyzing the trials to criterion, errors, and number of correct choices on the first-quarter trials by separate $3 \times 3 \times 2$ analyses of variance and Newman–Keuls Studentized range tests. To further analyze the data, serial position curves were compiled.

Trials to criterion. Mean trials to criterion and standard deviations for the 6 groups are shown in Table 2. A $3 \times 3 \times 2$ analysis of variance, with 3 types of instructions, 3 replications, and 2 association values, found a significant main effect of association value, $F\ (1, 72) = 6.64, p < .05$. This indicates that, as expected, the HA trigrams were learned in less trials ($\bar{X} = 8.24$) than were the LA trigrams ($\bar{X} = 10.31$). The main effect of instructions approached significance, $F\ (2, 72) = 2.74, .05 < p < .10$. This suggests that there is a tendency for P instructions to induce the most effective learning strategy and the fastest learning ($\bar{X} = 8.10$ trials), while the control instructions, where *S* must find his own strategy, resulted in the slowest learning ($\bar{X} = 10.40$ trials).

TABLE 1

Mean Reaction Time (in sec.) to Position Numeral and to Preceding Item for the Six Experimental Groups

Instructions	Trigram association: Low — Serial RT	Low — Position RT	High — Serial RT	High — Position RT
Serial	4.68(1)	3.89(2)	2.69(1)	2.98(2)
	3.11(2)	4.77(1)	2.59(2)	3.91(1)
Position	3.74(2)	4.13(1)	2.42(2)	2.80(1)
	3.58(1)	1.85(2)	3.04(1)	1.95(2)
Control	4.16(1)	2.91(2)	3.13(1)	4.25(2)
	3.50(2)	3.74(1)	3.78(2)	3.52(1)

Note. Numbers in parentheses indicate order of testing. Abbreviation: RT = reaction time.

TABLE 2

Mean Trials to Criterion (and Standard Deviations) for the Six Experimental Groups

Trigram association	Instructions: Serial	Position	Control
Low	11.13	7.87	11.93
	(4.12)	(2.05)	(5.80)
High	7.53	8.33	8.87
	(3.62)	(3.56)	(2.67)

In order to determine the significance of the specific group differences, an internal comparison of the cell means was made by the Newman–Keuls Studentized range test. This found only that Group C–LA took significantly more trials to learn the list than Group P–LA ($\bar{X} = 11.93$ and 7.87, respectively, $p < .05$). Thus, the main reason for the borderline significance of instructions is the difference between Groups C–LA and P–LA.

Errors. Mean errors (intrusions and omissions) and standard deviations for the 3 types of instructions and 2 association levels are presented in Table 3. The results of the $3 \times 3 \times 2$ analysis of variance of the error data found significance in the main effect of instructions, $F\ (2, 72) = 6.90, p < .025$; the Instructions × Association Value interaction $F\ (2, 72) = 4.12, p < .05$; and the Instructions × Lists interaction, $F\ (2, 72) = 2.82, p < .05$. Again, the significant main effect of association value

TABLE 3

MEAN ERRORS (AND STANDARD DEVIATIONS) FOR THE SIX EXPERIMENTAL GROUPS

Trigram association	Instructions		
	Serial	Position	Control
Low	63.20 (22.25)	38.87 (12.32)	65.40 (30.44)
High	37.60 (18.43)	44.33 (23.82)	50.07 (20.14)

indicates that, as expected, more errors were made while learning LA trigrams ($\bar{X}$ = 55.82) than HA trigrams ($\bar{X}$ = 44.00). That the main effect of instructions is significant indicates that the P groups made the fewest errors ($\bar{X}$ = 41.60), while the C group made the most ($\bar{X}$ = 57.74). Again S groups were intermediate ($\bar{X}$ = 50.40) between P and C groups. Overall, the P instructions induced the most efficient strategy (in terms of errors), while the control instructions, leaving *S* to search for a strategy by himself, were least efficient. The significant Instructions × Association Value interaction indicates that this was especially true for the LA trigrams, whereas for the HA trigrams, the S instructions induced a more effective strategy, resulting in fewer errors, than either of the other instructions. A Newman–Keuls Studentized range test was performed to determine the exact locus of the experimental effects. This analysis indicated that the significant Instructions × Association Value interaction was due to the fact that in the LA groups, the P instructions resulted in fewer errors than either S instructions or C instructions ($\bar{X}$ = 38.87, 63.20, and 65.40, respectively, $p < .05$), while in the HA groups there were no differential effects due to instructions. Thus, it seems that for LA trigrams, the P instructions induced the most efficient strategy, while for HA trigrams, the type of instructions had little effect.

Number correct on first-quarter trials. An analysis of the number of correct responses on trials during the first quarter of learning was suggested by results of an experiment by Postman and Stark (1967), in which evidence supporting the serial hypothesis was found when the first few trials were analyzed, but not in the analysis of the number of correct responses on the entire list. Mean numbers of correct trigrams and standard deviations for the first quarter of learning are found in Table 4. An analysis of variance found no significant differences among any of the means. The most probable reason for the discrepancy between these results and those found by Postman and Stark lies in the different procedures used. Postman and Stark used a PA-SL transfer procedure, while this experiment used a procedure which induced various learning strategies by means of different instructions. Apparently, any strategy induced by a PA task can be tapped early in serial learning, whereas the differential effectiveness of strategies suggested by varying instructions does not appear until later in learning.

Serial position curves. Serial position curves for HA and LA groups are shown in Figures 1 and 2 respectively. It can be seen from Figure 1 that the shape of the curve for Group S–HA is somewhat different from that normally encountered, in that there is no major drop in the percentage of errors at the end of the curve. While this is a rather sharp departure from the typical serial position curve, it is exactly what would be expected if *S*s learned the list by the serial instructions; that is, if each *S* made a connection between each pair of syllables, one would expect the beginning of the list to be learned most rapidly, with

TABLE 4

MEAN NUMBER CORRECT (AND STANDARD DEVIATIONS) ON THE FIRST-QUARTER TRIALS FOR THE SIX EXPERIMENTAL GROUPS

Trigram association	Instructions		
	Serial	Position	Control
Low	2.02 (1.15)	2.89 (1.37)	2.83 (.91)
High	3.14 (1.18)	2.78 (1.06)	2.70 (1.52)

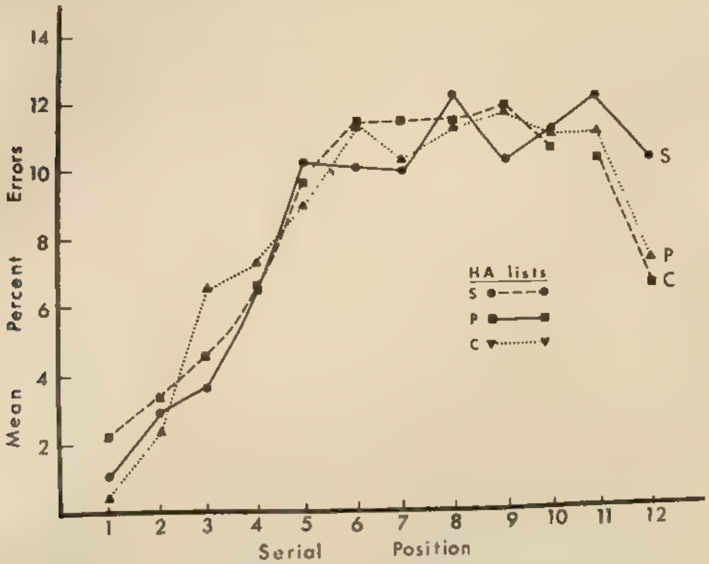


FIGURE 1. Serial position curves, showing percentage of errors for each serial position, for the 3 high-association (HA) groups. (Abbreviations: S = groups given serial instructions; P = groups given position instructions; and C = control groups.)

more and more difficulty occurring as *S* gets further and further into the list. The serial position curves for all low-association syllables, on the other hand, are more like the typical inverted-U-shaped serial position curves, with the ends showing fewer

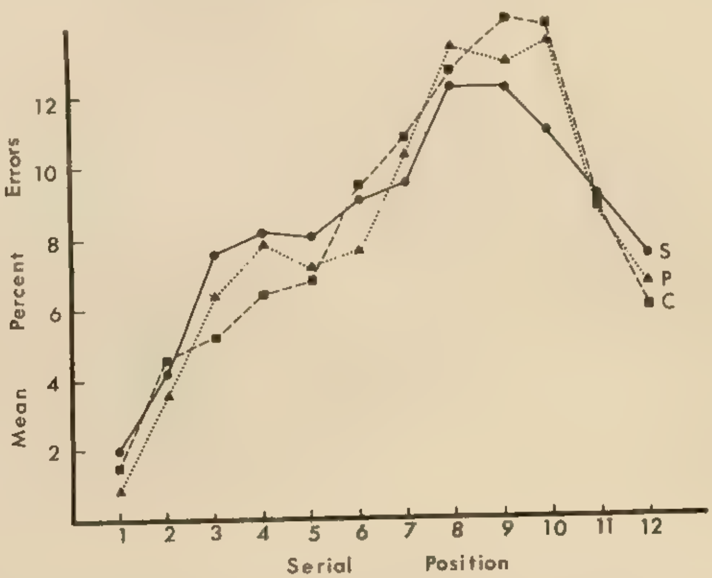


FIGURE 2. Serial position curves, showing percentage of errors for each serial position, for the 3 low-association (LA) groups. (Abbreviations: S = groups given serial instructions; P = groups given position instructions; and C = control groups.)

errors, and most errors occurring in the middle of the list.

Discussion

Two findings from this study seem worthy of further note. First is the effectiveness of instructions in influencing *S*'s learning strategy, especially for the error data. Most of the experimentation performed up to the present has been concerned with determining the strategy that *S*s prefer to use, rather than comparing the relative efficiency of various strategies (e.g., Brown & Rubin, 1967). The present experiment, on the other hand, is an attempt to compare the effectiveness of the specificity and position hypotheses when they are induced by instructions prior to learning. Although previous experiments have employed variations in instructions as an experimental variable, most have done so for other purposes, e.g., to contrast intentional (instructed) and incidental (noninstructed) learning (e.g., Dornbush & Winnick, 1967), to provide a basis for differences in motivation (e.g., Baron, 1967), or to provide information or misinformation about the learning situation (e.g., Holden & Rotter, 1962). The present experiment, by contrast, has provided instructions about specific procedures that are to be used to accomplish learning. That the instructions were acted upon by *S*s in this experiment seems to be indicated by the reaction time data, which showed that S groups built up stronger associations between one trigram and the next one on the list, and that P groups built up stronger associations between the number of the trigram on the list and the trigram than between successive trigrams. Furthermore, the significance of the instructional effect in the error data, as well as its borderline significance on trials to criterion, indicates that the specific suggestion of a strategy provides these *S*s with something of an advantage over control *S*s.

One possible explanation for these findings is that giving *S* instructions about how to learn the list allows him to start learning the list without having to search for a strategy. He would then learn the list more rapidly than if he had searched for a strategy, as long as the instructed strategy was not one which would hinder learning.

In the search for differences between the 2 instructed groups, the Newman–Keuls analysis found that Group P–LA made significantly fewer errors than either Group S–LA or Group C–LA. The latter 2 groups did not differ significantly from each other, indicating superiority of position instructions for the LA list. However, the Newman–Keuls analysis of the trials-to-criterion data showed that Group C–LA took significantly more trials to learn the list than Group P–LA, but found no significant difference between Group P–LA and Group S–LA. This is not meant to imply that *S* used only the strategy he was instructed to use. It is possible and not contradictory to the present results that some *S*s used some complex strategy based on the instructions.

The second finding of note is that the effectiveness of a learning strategy (as induced by instructions) is functionally related to the characteristics of the materials to be learned. That this was the case in the present experiment is indicated by the significant interaction between instructions and materials for the error data. This significant interaction revealed, first, that position instructions were more effective than specificity instructions with low-association trigrams, and second, that neither strategy had an advantage for high-association trigrams. This finding is quite understandable. Low-association trigrams, by their very nature, provide little basis for the interitem linkages demanded by the specificity strategy. The existence of numerical position in the list is therefore a feasible aid. As trigrams increase in association value, it should become easier to form interitem linkages, so that when high-association trigrams are used, the advantage of the position strategy is lost, and the strategies are of equal efficiency. It seems reasonable to predict that interitem linkages would be easier to form with words than with trigrams, and the position strategy should therefore have little advantage when words are used.

While the above reasoning is logical, at least one other possibility must be considered. It is feasible that, although *S*s were given instructions, and although reaction time data give a clue that those instructions were used, *S*s may have been using other cues in addition to and/or instead of the one suggested by the instructions.

The results from the present study indicate that the effectiveness of a particular strategy of learning may be functionally related to the characteristics of the materials to be learned. Caution suggests that there may also be a relationship to the method of learning employed. The present study, rather than using the more common method of anticipation, followed the suggestion of Battig and Lawrence

(1967) that the serial recall procedure is superior because it reduces the confusion induced by the anticipation method's "requirement that *S*s simultaneously try to learn each presented item and anticipate the following item . . . [p. 177]." Their conclusion that the recall procedure is more sensitive than the anticipation procedure for identifying the process in serial learning seems to be corroborated by the present results.

REFERENCES

Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74** (10, Whole No. 497).

Baron, R. A. Motivation and performance in skill and chance defined situation. *Psychonomic Science*, 1967, **9**, 631–632.

Battig, W. F., & Lawrence, P. A. The greater sensitivity of the serial recall than the anticipation procedure to variations in serial order. *Journal of Experimental Psychology*, 1967, **73**, 172–178.

Brown, S. C., & Rubin, E. D. Cue utilization in serial learning. *Journal of Experimental Psychology*, 1967, **74**, 550–555.

Dornbush, R. L., & Winnick, W. A. Short-term intentional and incidental learning. *Journal of Experimental Psychology*, 1967, **73**, 608–611.

Ebenholtz, S. M. Position mediated transfer between serial learning and a spatial discrimination task. *Journal of Experimental Psychology*, 1963, **65**, 603–608.

Holden, D. B., & Rotter, J. B. A nonverbal measure of extinction in skill and chance situations. *Journal of Experimental Psychology*, 1962, **63**, 519–520.

Postman, L., & Stark, K. Studies of learning to learn: IV. Transfer from serial to paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 339–353.

Winnick, W. A., & Dornbush, R. L. Ordinal position in serial learning. *Journal of Experimental Psychology*, 1968, **78**, 536–538.

Young, R. K. Tests of three hypotheses about the stimulus in serial learning. *Journal of Experimental Psychology*, 1962, **63**, 307–313.

Young, R. K. Serial learning. In T. D. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

(Received July 20, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 678–686

SEQUENCE OF RESPONSE DEVELOPMENT IN HUMAN EYELID CONDITIONING

ARCHIE B. LEVEY[1]
Clinical Psychiatry Unit, Medical Research Council, Graylingwell Hospital, Chichester, England

IRENE MARTIN
University of London

The acquisition of the conditioned eyelid response was examined fr standpoint of response patterning to determine whether respon develops in an orderly sequence independent of response frequenc analysis was confined to unreinforced trials in order to avoid maskin characteristic response shapes by the unconditioned response (UCR influence of response frequency was assessed by varying air-puff i and interstimulus interval. The 4 response shapes identified were (a conditioned response alone, (*b*) an unaccompanied response in the lat range of the UCR, (*c*) a double response combining *a* and *b*, and (*d*) a r in which *a* and *b* were indistinguishably blended. It was found th eyelid response exhibits an orderly sequence of development culmin the blended response, though this development is characterized by fr return to the earlier response types.

A lively interest in the analysis of specific response features and patterns of responding characterized the earlier experimental literature on conditioning (e.g., Bernstein, 1934; Hilgard, 1931; Humphreys, 1943; Marquis & Porter, 1939; Wendt, 1930). Attention was also focused during this period on the analysis of relationships between the conditioned response (CR) and the unconditioned response (UCR) (e.g., Grant, 1939; Hilgard, 1936; Miller & Cole, 1936). In subsequent years, the response itself tended to be taken for granted as interest in conditioning turned towards theory testing and the analysis of stimulus effects. The last decade, however, has seen a revival of interest in the description and analysis of response characteristics and of response changes during conditioning (e.g., Grevert & Moore, 1970; Grings & Schell, 1969; Kimble & Ost, 1961; Kimmel, 1966; Lockhart, 1972), including changes in overall response topography (e.g., Ebel & Prokasy, 1963; Gormezano, 1972; Grings, Lockhart, & Dameron, 1962; Prokasy, 1965; Prokasy, Ebel, & Thompson, 1963; Smith, 1968). At the same time, new measures have been dev d which describe topographic aspec the response with increasing accuracy (Edelberg, 1970, 1972; Pennypacker, 1964).

Objective measures of response efficiency have been reported for the conditioned eyelid response which provide a precise, quantitative description of the degree to which the developing CR anticipates the occurrence of the UCR (Levey & Martin, 1968; Martin & Levey, 1969). Investigation of these measures has shown that (*a*) they increase linearly toward a stable end point in the course of acquisition, (*b*) the increase is characterized by random fluctuations, (*c*) rate of increase is largely independent of response frequency, and (*d*) the end point may or may not involve optimal attenuation of the air puff, depending upon stimulus and *S* variables. The last finding is in accord with the observation that use of paraorbital shock produces efficient placing of the CR, though attenuation is not involved (Coleman & Gormezano, 1971). In the classical conditioning paradigm, response shaping is determined by *S* rather than by *E* (cf. Prokasy, 1965), and the increase in response efficiency may be regarded as a form of adaptive responding.

A tentative formulation that attempts to account for these findings has been

[1] Requests for reprints should be sent to Archie B. Levey, Medical Research Council, Clinical Psychiatry Unit, Graylingwell Hospital, Chichester, Sussex PO19 4PQ, England.

offered elsewhere (Martin & Levey, 1969, pp. 119–136) and will not be described in detail. Briefly, it is suggested that the conditioned stimulus (CS) and the unconditioned stimulus (UCS) are combined, from the beginning of acquisition, in an internally represented CS/UCS complex, to which the organism responds as a single stimulus entity, and for which a neuronal model is formed, analogous to Sokolov's (1963) model of the orienting stimulus. It is assumed that the earliest response to this stimulus complex is the formation of an internal response model. In early trials the stimulus complex evokes a UCR replica in the temporal range of the CS, which is incorporated in the response model and becomes the basis of the anticipatory CR. In the course of acquisition, the internal model of the response complex is progressively adjusted, through comparator mechanisms, to the combined characteristics of the stimulus complex. The point in acquisition at which an overt response is elicited depends on performance factors, and the response form is governed by the current state of the response model. The final response is an integrated complex that represents efficient use of all the information coded in the stimulus complex, and hence produces efficient responding.

The present study was undertaken to test some of the observable consequences of the formulation outlined above. (*a*) Presentation of the CS alone in early trials should be capable of eliciting an attenuated response in either the CR or the UCR latency range or both. (*b*) On intermediate trials the CS should elicit a combined response in which both CR and UCR components are identifiable. (*c*) On trials late in the acquisition series the CS should be able to elicit a response in which CR and UCR components are combined in a single entity. (*d*) The *S*s who respond late in the acquisition series, assuming response probability to be a function of various performance factors, should produce responses that reflect the development of the internal response model. A corollary of these inferences is that the developing response should exhibit an orderly sequence, independent of the level of response frequency.

The test of these propositions must rest with unpaired presentation of the CS, since occurrence of the UCS would mask the response shape by forcing a UCR. Accordingly, the study used the method of unreinforced test trials, though this carries the unavoidable limitation that the effect of partial reinforcement as a variable cannot be independently assessed. This procedure has the sanction of other recent investigations, however (e.g., Coleman & Gormezano, 1971), and in the present context the test trials can be regarded as a kind of "window" on the current state of the developing response complex.

Method

The *S*s were 72 naive adult males, screened for sensory defects, who were recruited as paid volunteers. The conditioning procedure involved *S*s being conditioned to a 1,000-cps 65-db. tone CS presented through headphones, and an air-puff UCS of 60-msec. duration, delivered to the cornea through 1-mm. tubing and terminated with CS termination. Stimulus presentations and durations were controlled by a Devices Digitimer. The eyelid response was recorded by means of an ORP-12 Cadmium Sulphide photocell whose output was fed, with the marker signals, to a 2-channel Elema Schönander Mingograph with a chart speed of 50 mm/sec. The *S*s were seated in a dimly illuminated cubicle, screened from the apparatus and from *E*. Thirty acquisition trials were given with 18 unreinforced test trials interspersed randomly throughout the series, with the limitation that no 2 test trials occur in succession. The same randomization order was used for every *S*. Immediately following the acquisition trials, extinction trials were run to a criterion of 5 successive nonresponse trials or a limit of 20 trials.

In order to manipulate response frequency, 2 experimental factors known to affect this variable were accounted for in the design. Half of the *S*s were conditioned at short (400 msec.) and half at long (800 msec.) CS durations, crossed with 2 levels of UCS intensity, weak (3 psi) and strong (6 psi) air puffs. Since personality variables, notably extraversion and neuroticism, are known to affect response frequency, *S*s were assigned to the 4 cells of the design on the basis of these 2 personality factors and their interaction, using pretest scores on the Eysenck Personality Inventory (Eysenck & Eysenck, 1964).

Responses were scored on each of the 18 test trials for all deflections of 1 mm. or more above the stable baseline. A response was scored as a CR if

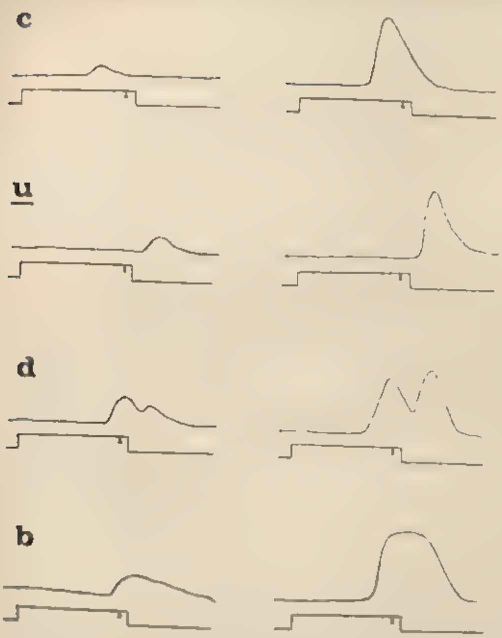


FIGURE 1. Chart tracings of the 4 response types discussed in text: conditioned response (CR) only (c), unreinforced responses falling in unconditioned response range (*UCR*) only (*u*), double CR/*UCR* (d), and blended CR/UCR (b). (The figures on the left represent early trials, while those on the right are taken from later trials. The marker indicates conditioned stimulus [CS] onset, duration, and offset. The arrow indicates the point at which the unconditioned stimulus would have occurred. The CS duration in all figures is 800 msec.)

it fell between an onset latency of 120 msec. and the offset of CS, and as a *UCR* if it occurred within 120 msec. following CS offset. (The italicized label will be used throughout this article to refer to unreinforced responses falling in the UCR latency range.) It should be emphasized that the latter responses, though defined by the UCR latency range, were not simply duplicates of the UCR, either in shape or amplitude. Many of them, particularly in earlier trials, were low in amplitude and did not show the characteristic sharp rise of the UCR to UCS presentation. It should also be noted that the findings of this study do not apply to voluntary-form responding, which did not occur in the sample studied.

Responses were then categorized into the 4 types illustrated in Figure 1. This categorization was done by a research assistant thoroughly familiar with conditioning records, but unaware of the purpose of the analysis.[2] The types identified were: CR only (C), *UCR* only (*U*), double response (D) and blended response (B). The categorization was objectively based on the latency criteria outlined and on the locus of inflection points in the wave forms in the case of double responses. On this basis a small number of multiple CRs was included in the CR-only category. Obvious spontaneous blinks, defined by their characteristic rise time, amplitude, and duration, were separately identified and excluded from the analysis. Coding reliability was assessed by independently rescoring one third of the records, resulting in a 95% overall agreement. While this type of categorization is crude, by comparison with the objective measures of response efficiency mentioned earlier, it served to identify the basic response shapes with which the study was concerned.

The sequence of response types was identified for each *S* by the following method: each first appearance of any response type was recorded for every *S* in order of occurrence. This yielded a code containing 4 successive elements. For example, if *S* gave the following sequence: *U*, C, C, *U*, D, *U*, D, C, C, B, C, D, it was coded as *U*-C-D-B, indicating that the progression of response types occurred in that order. This method of summary yielded 24 possible sequences representing the permutations of 4 elements. An appreciable number of *S*s showed only 3 of the response types. In this situation, it follows that the fourth element of the code can only be the remaining type, and these codes were therefore completed with the missing element. A few *S*s gave only 2 of the response types, and were randomly assigned to those codes in which the 2 relevant response types appeared first in the order observed. It is important to note that this method of coding takes no account either of the absolute number of responses of a given type, or of the order in which any of the response types, having first appeared, may recur prior to the emergence of the next type. The main hypothesis to be tested was that the 4 types occur in ordered sequence. A sequence such as C-*U*-D-B, should occur more frequently, for example, than the sequence D-*U*-C-B, while the sequence B-D-*U*-C should be unlikely. In other words, it was postulated that the developmental sequence begins with the single responses, progresses through the double response, and ends with the integrated blend response.

RESULTS

The total number of responses categorized was 499, divided into 240 C, 91 *U*, 79 D, and 87 B responses. The observed base level of responding immediately prior to the acquisition series was .069 for the shorter interval and .120 for the longer. These are average values based on 3 samples from each *S*, allowing for the 120-msec. *UCR* criterion, and are within the ranges usually observed in our laboratories.

[2] The authors are most grateful to Mary Acock, Medical Research Council, for her cheerful and competent performance of this task.

Only one double response occurred and this is consistent with chance expectation at the observed levels. Analysis of variance on the 2 × 2 sampling model of the design, summing over 18 trials, showed that total response frequency was significantly lower under the interaction of a short interstimulus interval and a weak air puff, $F\ (1, 68) = 4.19,\ p < .05$. Separate analyses showed the *U* response to be slightly more frequent under the stronger air puff, $F\ (1, 68) = 4.98,\ p < .05$, while the D response was significantly more frequent with the longer interval, $F\ (1, 68) = 4.02,\ p < .05$. The C and B responses were not differentially affected by experimental conditions. Twelve *S*s failed to produce responses to the unreinforced trials and were excluded from further analysis.

The fact that the 4 response types could be identified in substantial numbers lends immediate support to the notions discussed in the previous section. These results indicate that the unpaired CS is capable of eliciting responses in both the CR and UCR latency ranges, singly and in combination, including a completely blended composite of both responses. The question of interest, however, is the order in which these response types occurred. If they merely represent 4 discrete response shapes or patterns, it might be expected that each of the sequences would be equally probable. In fact, of the 24 possible codes, the 2 expected sequences, C-*U*-D-B and *U*-C-D-B, accounted for over 50% of all responders.

The assumption that each of the sequences is equally probable is open to several objections. In particular, any factors that govern the proportions of the response types would also dictate the relative frequencies of the various sequences. Thus, if chance alone determined the first response type, the predominance of C, noted earlier, would lead to a higher frequency of all codes beginning with C. Using the same argument for each position of the elements in the codes, the most probable sequence would be C-*U*-B-D, and the next most probable would be C-*U*-D-B, representing a spurious advantage to the sequence expected by the theoretical model. In the following analyses, therefore, the calculation of expected frequencies for each of the sequence codes used the joint probabilities of the response types, based on their overall frequencies, assuming independence. This afforded an effective control for base-level responding in relation to the asymmetric latency criteria for the C and *U* responses, for experimental factors, and for any unknown factors which may have determined the distribution of response likelihoods.

No hypothesis was offered to predict whether the CR would precede the *UCR*, or vice versa. In order to summarize the data, therefore, and reduce the number of codes to manageable proportions, the CR and *UCR* responses were considered separately. This yielded 6 sequence codes based on the permutations of 3 elements. It should be noted that this procedure offered the further advantage of accounting for all *S*s, since for those few *S*s who gave only 2 response types, the code could be completed with the missing element.

Table 1 summarizes the data for 3 analyses: CR omitting *UCR*, *UCR* omitting CR, and CR combined with *UCR*. The symbol C*U* in the code label on the left indicates the position of these elements in the sequence, while the symbol at the head of each pair of columns designates the element itself. Thus, the first entry in the top row of the table is the number of *S*s who gave the sequence C-D-B, the second is the expected frequency for that sequence, the third is the number producing

TABLE 1

Observed (O) and Expected (E) Frequencies of *S*s Responding within Six Sequence Codes

Code	C		U		CU	
	O	E	O	E	O	E
C*U*-D-B	42	16.94	34	10.11	44	19.05
C*U*-B-D	6	18.64	6	11.13	8	20.97
D-C*U*-B	6	8.57	7	9.41	5	7.52
D-B-C*U*	3	3.07	10	9.01	2	1.96
B-C*U*-D	2	9.64	1	10.88	1	8.48
B-D-C*U*	1	3.14	2	9.46	0	2.02

Note. See text for code explanation.

the sequence *U*-D-B, and so on. With one exception, to be discussed later, the experimental conditions had no important effect on the sequences and will not be considered separately in these analyses.

Inspection of the table shows that the predicted sequences, C-D-B and *U*-D-B, were overwhelmingly highest in proportion, while B-D-C and B-D-*U* were scarceley represented, χ^2 (5) = 53.95 and 74.40, respectively, $p < .001$. The joint function of CR and *UCR*, shown in the third pair of columns, follows the same pattern, χ^2 (5) = 50.16, $p < .001$, with greater symmetry. Further clarification of the relationship between CR and *UCR* was afforded by examining the first response of each *S*. The observed frequencies of the response types C, *U*, D, and B were 36, 16, 7, and 1, respectively, and yielded χ^2 (3) = 13.24, $p < .005$. Interestingly, the observed frequencies of the CR and *UCR*, ignoring the remaining categories, lie close to the observed proportions of the 2 responses overall, suggesting that the CR is more likely as a first response simply because it is more available.

The combination of a short interstimulus interval and weak UCS not only reduced the level of responding, but virtually excluded the unexpected response sequences. For example, 90% of the *S*s responding under these conditions showed the sequence C-D-B. In order to explore this unexpected finding, the 60 responding *S*s were divided into 2 equal groups of 30 early responders and 30 late responders, both across experimental conditions and separately within experimental co[illegible]ons, and the data were reanalyzed [illegible]ese analyses showed the predicted [illegible]onse sequences to be undifferentiated b[illegible]arly or late responding. By contra[illegible] the deviant sequences were virtually [illegible]ted to late responders and in the case [illegible] the within-conditions comparison, enti[illegible]so. This result is consistent with 2 expl[illegible]ory factors. Where the experimental [illegible]ndi-tions are such as to suppress respo[illegible]ding, late responses represent the beginn[illegible]g of response development. Where the [illegible]onditions facilitate responding, late resp[illegible]ding represents those *S*s who withold res[illegible]onses until late in their own developmen[illegible]l sequence, but whose patterns of responding show an internal sequence of development.

No hypothesis was offered as to the behavior of response sequences du[illegible] extinction. The following analysis [illegible] empirical and is offered merely as a description of the typical sequences. It should be noted that the categorization of responses is made difficult du[illegible] because the response complex [illegible] to shift on its time base, and this ma[illegible]lt in misclassification of some resp[illegible]. Fifty-four *S*s produced a total of [illegible]11 responses in extinction, of which 235 were either blended or double responses. In these trials, 45% of all *S*s responding gave response sequences that fell in those categories beginning with the blended response. The next most frequent category comprised those sequences beginning with the double response. These results suggest that the response complex is in some sense breaking down, though not necessarily regressing in an orderly fashion through the reverse of the response-type sequences. The results must be viewed with some caution, since the appearance and disappearance of responses late in the sequence may produce artifacts that seem to resemble disintegration. It is noteworthy, however, that those *S*s giving a large number of responses in extinction were those more likely to follow the graded sequences. Table 2 shows the distribution of response sequences for the 25 *S*s giving either 3 or 4 response types

TABLE 2

FREQUENCIES OF RESPONSE SEQUENCES FOR 25 *S*s GIVING THREE OF FOUR RESPONSE TYPES IN EXTINCTION, COMBINING CR AND *UCR* ONLY RESPONSES

Response	3	4	Total	Expected
C*U*-D-B	5	0	5	3.62
C*U*-B-D	0	0	0	6.76
D-C*U*-B	2	2	4	2.65
B-C*U*-D	4	0	4	6.38
D-B-C*U*	3	1	4	2.45
B-D-C*U*	3	5	8	3.14
Total	17	8	25	

Note. See text for code explanation.

in extinction, combining CR and *UCR* in one category. Expected frequencies were calculated on the basis of overall response proportions in extinction. The average number of extinction trials for this group was 17.72. It is evident that the reversed sequences B-D-C*U* and D-B-C*U* account for a disproportionately high number of responders, χ^2 (5) = 17.36, $p < .005$. This suggests the intriguing possibility that if *S*s go on responding long enough in extinction, the response complex tends to be dismantled in reverse order to its development.

Discussion

To summarize the results of this study briefly, it has been shown that the emergence of response types categorized as CR, *UCR*, double CR/*UCR*, and blended or integrated CR/*UCR* occurs for the most part during acquisition in an orderly sequence of response development. The sequence as such is independent of the experimental conditions examined, although the absolute numbers of a given response category are affected. Those experimental conditions that suppress responding appear to delay the onset of the sequence, while those that facilitate responding show evidence that internal response development occurs early in acquisition. The sequence does not exclude a return to earlier modes of responding and is probably best described as an overall emergent trend in a setting of ongoing random variation consistent with the fluctuation of efficiency levels mentioned earlier. However, in the present study, which was limited to partial reinforcement, it is possible that extinction processes also contributed to the return to earlier response types.

These results were anticipated in a surprisingly neglected early study by Grant (1939), though differences in experimental conditions, procedures, and methods of recording make it impossible to compare the 2 studies directly. Grant identified 6 response patterns, of which 2, labeled the "unimodal CR" and the "completed slope CR," resemble the responses labeled double and blended responses in the present study. He noted of the latter response that "this very efficient pattern generally occurs later in the conditioning series [p. 452]," and commented both on the implied sequence of patterns and on the tendency toward "regressive shifts to patterns which had been present in an earlier stage of the experiment [p. 460]." Grant's study, of which the authors were unaware at the time the present study was designed and analyzed, was specifically aimed at describing response patterns and their shifts in relation to the question of CR/*UCR* relationships. Though the measurements were made entirely on reinforced trials, the similarity of results is striking.

The view that response efficiency is independent of response frequency is consistent with the recent finding of Coleman and Gormezano (1971) that the precision and variability with which the CR peak is placed in relation to UCS onset are uncorrelated with response frequency. In the present study, CR frequency accounted on the average for only one quarter of the variance associated with the 4 response types, and, in general, correlations among response types were nonsignificant. Nevertheless, the finding that response suppression may delay the onset of response development requires some modification of this view.

The development by Prokasy (1972) of a two-phase model that successfully predicts acquisition curves in human eyelid conditioning may offer a lead. One of the parameters (K) of this model reflects a factor of delay during early acquisition trials which Prokasy (p. 143) attributes in part to variables that affect orienting behavior. It seems entirely possible that the low saliency of the weak UCS and short interstimulus interval in the present study may have resulted in failure to attend during the early trials. It is reasonable to suppose that if the stimulus processing mechanism is "idling" during this period, the postulated stimulus and response models will be delayed in formation.

The interest of the present results, apart from their general support of the theoretical model outlined earlier, centers on the eyelid response at UCR latency. The occurrence of 170 of these responses, either alone or in combination with the CR, is well above the base-rate expectation of 28.4 responses and suggests that they are not trivial. A similar response, occurring on unreinforced trials, has been described by Zeiner (1968) for several autonomic measures (galvanic skin response, heart rate, salivation, and pupil dilatation). In electrodermal conditioning, it is commonly referred to as the second interval response (SIR), first described by Grings et al. (1962). Several explanatory concepts were examined by these authors, viz., response substitution, inhibition of delay, temporal conditioning, CS

offset response, and disparity responding. Reasons were given to show that without modification, none of these explanations is entirely satisfactory. Part of the interest of identifying the *UCR* or SIR in eyelid conditioning is the fact that the orienting component, which is prominent in the electrodermal system, can be largely ignored. The behavior of the alpha blink, or CS reflex blink, has been shown to resemble that of the orienting response (Levey & Martin, 1967; McAllister, 1953). This blink occurs typically to onset of the CS, but rarely to offset at the stimulus intervals employed in the present study. (We have informally observed CS-offset alpha blinks, though only to tone stimuli of very long duration, e.g., 120 sec.) Thus, the mechanisms of CS offset and disparity responding do not apply to the eyelid response.

Zeiner (1968) has shown empirically, in an elegant experiment comparing a trace paradigm in which the CS extended beyond the UCS, with a delay paradigm, that the SIR in electrodermal conditioning is not merely a response either to offset of the CS or to contiguity between CS offset and UCS onset. He has also presented cogent reasons, which are equally applicable to the present results, suggesting that the SIR is not due either to simple response substitution or to inhibition of delay. His conclusion was that the SIR phenomenon represents one part of a dual process, in which the occurrence of 2 separate responses is attributed to separate mechanisms of preparatory responding, mediated by learned fear, and a modified form of stimulus substitution. It is our view that in eyelid conditioning at least, a two-process model does not readily account for the eventual blending and integration of the 2 responses, CR and *UCR*, and the present results suggest that both are intrinsic components of the classical conditioning process.

Indeed, the most interesting aspect of the response sequences described is not the nature of the 2 response components but the fact that they combine into an integrated whole. The fact that a blend of CR and UCR occurs in eyelid conditioning is not in doubt (Spence, 1956). What the present study shows is that the blended response, as well as the composite double response, is elicited by the CS in the absence of the UCS. This of itself indicates that they are in some sense integrated into a single response. It is important to note that the blended response does not necessarily result in optimum avoidance of the air puff. Doubt has recently been thrown on the ability of simple reinforcement mechanisms to account for response placement (Gormezano, 1972) as well as response probability (Prokasy, 1972). The view proposed here attributes response integration to the modeling of the CR/*UCR* response complex on the CS/UCS stimulus complex, and this is consistent with our present knowledge, though it cannot be proven.

If the CR and *UCR* are equally parts of the classical conditioning process, there remains the consideration of why the *UCR* in the present study should have been less frequent than the CR. The probability of occurrence on the CR or the *UCR* as a first response has been shown to be entirely consistent with its overall probability. It has also been shown that with a stronger UCS the UCR occurs more frequently, indicating that the CS is actively conveying this information. If, as suggested, the CS and UCS form a single stimulus complex, sharing the information content of both, the ability of the CS to elicit the *UCR*, or the UCS to elicit the CR, might be expected to be weaker than when the 2 stimuli are presented together. The second half of this proposition cannot be tested, since the UCS alone cannot give rise to an *observable* CR. In the absence of this unobtainable evidence, the difference in frequency of the CR and UCR offers no basis for rejection of the conclusion that the CR and the *UCR* are equally parts of the classical conditioning sequence, and that their integration into the double or composite response and finally into a blended response is an intrinsic part of the conditioning process.

Finally, though the results lend support to the theoretical model outlined, they also raise a very puzzling question. The data clearly show that on very early trials the CS is capable of eliciting a CR or a *UCR*, but not both. It might be noted that this situation is equally problematic for a two-process model. Nothing in the proposed model or in any current theory of classical conditioning predicts this phenomenon. It is not the purpose of this article to enter into discursive theorizing. The data seem to suggest 3 possibilities. First, the early trials may recruit insufficient response energy, in some sense, to support 2 responses. In this event, the *UCR* would be given only on a trial on which the CR had for some reason not occurred. Second, if 2 separate processes are involved, they may for unknown reasons be mutually exclusive or may interfere in the

early trials. Third, and most attractive in our view, the phenomenon can be handled within the confines of the proposed model. If it is assumed, following Estes (1950), that the comparator process mentioned earlier, by which it was suggested the integrated response complex is gradually built up, operates on each trial by scanning only a random subset of the information in the stimulus complex, the result is at least partially explained. If this were the case, the occurrence of a CR or a *UCR* or of no response in early trials would be a function of the particular subset sampled. On successive trials, further information would lead to progressive shaping and integration of the complete response model. This view is consistent with current views of information processing and is able to account for the gradual integration of the response.

No final answer can be given to the problems raised by the occurrence of the *UCR* and the CR/*UCR* complex in response to the CS. While, at this stage, the results lend support to the view of conditioning outlined earlier, they do not, of course, confirm it. The purpose of the present study was to establish firmly the empirical basis of response development sequences in eyelid conditioning. Further work will be required to elucidate the underlying mechanisms and to arrive at a satisfactory explanatory formula.

REFERENCES

Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *Journal of General Psychology*, 1934, **10**, 173–197.

Coleman, S. R., & Gormezano, I. Classical conditioning of the rabbit's (*Oryctolagus cuniculus*) nictitating membrane response under symmetrical CS–US interval shifts. *Journal of Comparative and Physiological Psychology*, 1971, **77**, 447–455.

Ebel, H. C., & Prokasy, W. F. Classical eyelid conditioning as a function of sustained and shifted interstimulus intervals. *Journal of Experimental Psychology*, 1963, **65**, 52–58.

Edelberg, R. The information content of the recovery limb of the electrodermal response. *Psychophysiology*, 1970, **7**, 145–147.

Edelberg, R. Electrodermal recovery rate, goal orientation and aversion. *Psychophysiology*, 1972, **9**, 512–520.

Estes, W. K. Toward a statistical theory of learning. *Psychological Review*, 1950, **57**, 94–107.

Eysenck, H. J., & Eysenck, S. B. G. *Manual of the Eysenck Personality Inventory.* London: University of London Press, 1964.

Gormezano, I. Investigation of defense and reward conditioning in the rabbit. In A. M. Black & W. F. Prokasy (Eds.), *Classical conditioning II. Current research and theory.* New York: Appleton-Century-Crofts, 1972.

Grant, D. A. A study of patterning in the conditioned eyelid response. *Journal of Experimental Psychology*, 1939, **25**, 445–460.

Grevert, P., & Moore, J. W. The effects of unpaired US presentations on conditioning of the rabbit's nictitating membrane response: Consolidation or contingency. *Psychonomic Science*, 1970, **20**, 177–179.

Grings, W. W., Lockhart, R. A., & Dameron, L. E. Conditioning autonomic responses of mentally subnormal individuals. *Psychological Monographs*, 1962, **76**(39, Whole No. 558).

Grings, W. W., & Schell, A. M. Magnitude of electrodermal response to a standard stimulus as a function of intensity and proximity of a prior stimulus. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 77–82.

Hilgard, E. R. Conditioned eyelid reactions to a light stimulus based on the reflex wink to sound. *Psychological Monographs*, 1931, **41**(1, Whole No. 184).

Hilgard, E. R. The nature of the conditioned response: I. The case for and against stimulus substitution. *Psychological Review*, 1936, **43**, 366–385.

Humphreys, L. G. Measures of strength of conditioned eyelid responses. *Journal of General Psychology*, 1943, **29**, 101–111.

Kimble, G. A., & Ost, J. W. P. A conditioned inhibitory process in eyelid conditioning. *Journal of Experimental Psychology*, 1961, **61**, 150–156.

Kimmel, M. D. Inhibition of the unconditioned response in classical conditioning. *Psychological Review*, 1966, **73**, 232–240.

Levey, A. B., & Martin, I. Reflex sensitivity and the conditioned eyelid response. *Psychonomic Science*, 1967, **8**, 153–154.

Levey, A. B., & Martin, I. The shape of the conditioned eyelid response. *Psychological Review*, 1968, **75**, 398–408.

Lockhart, R. A. Interrelations between amplitude, latency, rise time, and the Edelberg recovery measure of the galvanic skin response. *Psychophysiology*, 1972, **9**, 437–442.

Lockhead, G. R. Processing dimensional stimuli. *Psychological Review*, 1972, **79**, 410–419.

Marquis, D. G., & Porter, J. M. Differential characteristics of conditioned eyelid responses established by reflex and voluntary reinforcement. *Journal of Experimental Psychology*, 1939, **24**, 347–365.

Martin, I., & Levey, A. B. *The genesis of the classical conditioned response.* Oxford: Pergamon Press, 1969.

McAllister, W. R. Adaptation of the original response to a conditioned stimulus. *Proceedings of the Iowa Academy of Science*, 1953, **60**, 534–539.

Miller, J., & Cole, L. E. The influence of a "voluntary" reaction upon the development and the extinction of the conditioned eyelid reaction. *Journal of Genetic Psychology*, 1936, **48**, 405–440.

NORMAN, D. A. (Ed.) *Models of human memory.* London: Academic Press, 1970.

PENNYPACKER, H. S. Measurement of the conditioned eyelid reflex. *Science,* 1964, 144, 1248-1249.

PROKASY, W. F. Classical eyelid conditioning: Experimenter operations, task demands and response shaping. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1965.

PROKASY, W. F. Developments with the two-phase model applied to human eyelid conditioning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory.* New York: Appleton-Century-Crofts, 1972.

PROKASY, W. F., EBEL, H. C., & THOMPSON, D. D. Response shaping at long interstimulus intervals in classical eyelid conditioning. *Journal of Experimental Psychology,* 1963, 66, 138-141.

SMITH, M. C. CS-US in … intensity i… classical conditioning o… nictitati… membrane response. *Jo… parative an… Physiological Psychology,* 1968, 66 …9-687.

SOKOLOV, E. N. *Perception and th… tioned refle…* Trans. by S. W. Wayd… Oxford, England Pergamon Press, 1963.

SPENCE, K. W. *Behavio… nd conditionin…* New Haven: Yale Univ… s. 1956.

WENDT, G. R. An analy… of the condi tioned knee-jerk. *Arch… ology,* 1930 19(No. 123).

ZEINER, A. R. Second int… nation conditioning of the GSR a… of UCS intensity and trace and d… oning paradigms. *Journal of Exper… ology,* 1968, 78, 267-280.

(Received Jul…

Manuscripts Accepted for Publication in the

Journal of Experimental Psychology

Conservation of Weight in Self, Others, and Objects. George B. Macready (Department of Measurement and Statistics, University of Maryland, College Park, Maryland 20742) and Corise Macready.

The Interaction of Drive Level and Task Complexity in Verbal Discrimination Learning. Jeffrey A. Seybert (Department of Psychology, University of Oklahoma, 455 West Lindsey, Room 705, Norman, Okahoma 73069), Dan M. Wrather, N. Jack Kanak, and Ed Eckert.

Affective Arousal and Energization Properties of Positive and Negative Stimuli. John Lamberth (Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122) and John Davis.

Proactive Interference in Short-Term Retention and the Measurement of Degree of Learning: A New Technique. John J. Shaughnessy (Department of Psychology, Northwestern University, Evanston, Illinois 60201), Ronald H. Nowaczyk, and Joel Zimmerman.

Stimulus Encoding in A-Br Transfer. John H. Mueller (Department of Psychology, University of Missouri-Columbia, 209 McAlester Hall, Columbia, Missouri 65201), Prentice Gautt, and James H. Evans.

Form Class as an Effective Encoding Dimension in Short-Term Memory. Charles P. Bird (Department of Psychology, Florida State University, Tallahassee, Florida 32306) and C. James Goodwin.

Sensory and Verbal Coding Strategies in Subjects with Absolute Pitch. Jane A. Siegel (Department of Psychology, University of Western Ontario, London, Ontario N6A 3K7, Canada).

The Backward Masking of Conditioned Stimuli: Effects on Differential and Single Cue Classical Conditioning Performance. Leonard E. Ross (Department of Psychology, University of Wisconsin, Psychology Building, Charter at Johnson, Madison, Wisconsin 53706), M. Cecilia Ferreira, and Susan M. Ross.

Retrieving Attribute and Name Information from Semantic Memory. Elizabeth F. Loftus (Department of Psychology, University of Washington, Seattle, Washington 98105) and William Cole.

The Temporal Course of Recovery from Interference and Degree of Learning in the Brown-Peterson Paradigm. Delos D. Wickens (Department of Psychology, Ohio State University, 321 Arps Hall, 1945 North High Street, Columbus, Ohio 43210) and M. M. Gittis.

The Effects of Mnemonic Learning Strategies on Transfer, Interference, and 48-Hour Retention. Douglas H. Lowry (Department of Social Sciences, Michigan Technological University, Houghton, Michigan 49931).

Structural Effects of Identical Elements among Stimuli with Meaningful Units. Willard N. Runquist (Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E1, Canada).

Overt Rehearsal and Long-Term Retention. Gary F. Meunier (Department of General and Experimental Psychology, Ball State University, Muncie, Indiana 47306), Jane Kestner, Jo Ann Meunier, and Douglas Ritz.

(*Continued on page 695*)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 687–690

HAPTIC EQUIVALENCE MATCHING OF CURVATURE BY BLIND AND SIGHTED HUMANS[1]

PHILIP W. DAVIDSON[2] AND TERESA T. WHITSON
Washington College

The relations between active handling, accurate haptic curvature matching, and task difficulty were studied using increases in the number of stimuli contained in comparison arrays to vary task difficulty. Blind *S*s were more accurate than sighted *S*s as task difficulty increased and marked differences in exploratory search were noted between blind and sighted *S*s. Moreover, the blind's characteristic search strategy, infrequently used by the sighted, led to improved matching by sighted *S*s restricted to its use. Also, in the larger Comparison-Array conditions, the more stimuli in the comparison arrays the more similar was the scanning of sighted *S*s in each condition to that of the blind. It was concluded that an interaction between task requirements and search style may in part determine accuracy in judgments of form properties like curvature.

Skill in haptically judging form properties such as length, proportion, or curvature seems to depend upon the particular pattern of active finger, hand, or arm movements used to explore the stimulus. However, little is known of the role of such movements in perception.

A recent study by Davidson (1972a) attacked one aspect of this question, the relation between the stimulus property being judged and the style of search. The experiment compared blind and sighted *S*s in discriminating curved stimuli by hand. The results showed that blind *S*s (*a*) were more accurate than the sighted with the most difficult curves, and (*b*) employed a scanning strategy infrequently used by the sighted, one leading to improved judgments by sighted *S*s restricted to its use.

The blind's strategy, called gripping, entailed using three or four fingers to simultaneously explore numerous points along the curve. Sighted strategies, although similar to gripping in hand and arm motion, involved successive search through pinching, or sweeping, the curve with one or two fingers. Davidson (1972a) concluded that gripping probably was suitable for prehending curves of low discriminability since it provided more stimulus information than other methods.

The present experiment examined another related aspect of this question, asking if the effectiveness of search strategies is influenced by the difficulty of the task itself, apart from the nature of the stimulus attribute. We used curvature equivalence matching, varying task difficulty by systematically changing the number of alternatives to be compared to the standard. The larger the number of stimuli in the comparison array, the greater the burden placed on immediate recognition memory and the more difficult the task (Goodnow, 1971). Part I compared the unrestricted exploratory search of blind and sighted *S*s while Part II tested the effectiveness of these search strategies.

METHOD

Subjects. Twenty-four *S*s with normal vision (M_{age} = 15 yr., 5 mo.) and 23 congenitally totally

[1] This research was supported by U.S. Public Health Service Research Grant 1-R01-EY-00807-01 from the National Eye Institute to the first author. Preparation of the manuscript was supported in part by grants to the University of North Carolina from the Maternal and Child Health Services (Project 916 to the Division for Disorders of Development and Learning) and the U.S. Public Health Service (Grant HD-03110 to the Biological Science Research Center). The authors gratefully acknowledge the cooperation of the Maryland School for the Blind and the Kent County High School. We also wish to thank M. Belcher and M. Logue for scoring the videotaped data, and E. Lehman for helpful suggestions for data analysis.

[2] Requests for reprints should be sent to Philip W. Davidson, Division for Disorders of Development and Learning, Child Development Institute, University of North Carolina, Chapel Hill, North Carolina 27514.

blind *Ss* (M_{age} = 19 yr., 0 mo.) participated in Part I. The blind and sighted *Ss* were matched approximately for IQ based on Verbal Scale scores on the Wechsler Intelligence Scale for Children (M_{IQ} = 102, SD = 9). Ninety paid undergraduate volunteers participated in Part II.

Stimuli. Seven different concave curves (as illustrated in Davidson, 1972a) were used, with arc heights of 2, 4, 6, 8, 10, 12, and 14 mm.

The task. The blindfolded *S* felt first the standard stimulus for 8 sec., then successively presented comparison stimuli for 8 sec. each to match the standard. The 2-, 4-, and 6-mm. curves served as standards (the greatest differential scanning effects reported by Davidson, 1972a, occurred with these three stimuli), each appearing 8 times in a random order for a total of 24 trials.

Task difficulty was varied by changing the number of curves appearing in comparison arrays and the same three difficulty conditions were used in both Parts I and II. In the first condition, each comparison array contained one curve (i.e., a *same–different* task), in the second, comparison arrays contained three curves (the standard plus two randomly selected others), and in the third, comparison arrays contained five curves (the standard plus four randomly selected others). The standard and comparison stimulus were the same 50% of the time in the One-Comparison condition and stimulus combinations and orders were randomized for all Comparison-Array conditions. No stimulus appeared twice in the same comparison array. Stimuli were presented following Davidson's (1972a) procedure.

Part I procedure. Seven blind and eight sighted *Ss* were tested in the One-Comparison condition, and groups of eight blind and eight sighted *Ss* were tested in the Three- and Five-Comparison conditions. Each *S*'s exploratory activity was videotape recorded and the recordings analyzed by two independent scorers following Davidson's earlier procedures (1972a). The scorers categorized an *S*'s scanning each time he felt a curve, noting only whether particular predefined methods occurred. The tapes were scored for gripping, pinching, and top sweeping, the three scanning strategies differentiating blind and sighted *Ss* in Davidson's (1972a) study and defined and illustrated there.

Part II procedure. In each of the three task conditions, 10 *Ss* were restricted to using only gripping, 10 *Ss*, to only pinching, and 10 *Ss*, only top sweeping. Each *S* was trained in using the appropriate scanning method before testing began.

RESULTS

Part I matching errors. The upper part of Table 1 gives the mean error scores from Part I. A two-way analysis of variance revealed effects of Vision, F (2, 41) = 11.71, $p < .01$, and Comparison Array Size, F (2, 41) = 48.78, $p < .01$. Subsequent t tests indicated that blind and sighted *Ss*' errors did not differ in the One-Comparison condition, t (41) = 1.3[illegible], $p > .05$, but that the blind made fewer errors in the Three-Comparison, t (41) = 2.46, $p < .01$, and the Five-Comparison condition, t (41) = 1.99, $p < .05$. Newman-Keuls tests showed that errors increased for both blind and sighted *Ss* with each increase in comparison array size (p < .01 for all comparisons). No change in errors occurred over trials, and IQ had no influence on errors, regardless of Vision or Comparison-Array condition.

Part II matching errors. The mean error scores from Part II are shown in the lower portion of Table 1. A two-way analysis of variance revealed effects of both Scanning, F (2, 81) = 4.03, $p < .01$, and Comparison Array Size, F (2, 81) = 53.31, $p < .01$. Newman-Keuls tests revealed that (*a*) gripping led to fewer errors than both pinching, $p < .01$, and top sweeping, $p < .01$, but pinching and top sweeping did not differ from one another, $p > .05$, and (*b*) errors increased with successive increases in comparison array size (p < .01 for all comparisons).

Comparing the Grip groups with the combined errors of the Pinch and Top Sweep groups showed that the three strategies led to similar errors in both the One-Comparison, t (28) = 1.31, $p > .05$, and the Three-Comparison condition, t (28) = 1.07, $p > .05$. However, gripping led to fewer errors than other methods in the Five-Comparison condition, t (28) = 2.36, $p < .02$. This outcome departs from our findings in Part I where, with scanning free to vary, the between-groups difference in the Five-Comparison condition was somewhat less than the Three-Comparison difference.

Analysis of scanning in Part I. After a check for interscorer agreement (the scorers agreed 83.9% of the time), following Davidson's (1972a) procedure we calculated the percentage of an *S*'s judgments in which the scorers reported seeing each scanning strategy and then obtained group averages. The resulting distributions for blind and sighted *Ss*, shown in Figure 1, were compared with separate Mann-Whit-

TABLE 1

Mean Curvature Matching Errors

Group	Number of Stimuli in Comparison Arrays					
	One		Three		Five	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Part I: exploratory scanning unrestricted						
Blind	6.57	1.76	10.37	2.44	15.62	1.99
Sighted	8.62	2.68	13.61	2.68	18.25	2.94
Part II: exploratory scanning restricted						
Grip	8.20	2.40	10.90	2.91	14.40	2.61
Pinch	9.40	1.74	11.40	2.76	16.60	2.28
Top sweep	9.40	2.57	12.80	2.48	17.00	2.64

ney *U* tests for each scanning strategy within each Comparison-Array condition.

In the One-Comparison condition, sighted *S*s pinched more often than the blind, U (7, 8) = 7.0, $p < .01$, while the blind *S*s gripped more often than the sighted, U (7, 8) = 13.0, $p < .05$, results paralleling Davidson's (1972a) previous study. The outcome with larger comparison arrays, however, differs from earlier findings for sighted *S*s. In the Three-Comparison condition, sighted *S*s' gripping showed a large increase compared to One-Comparison condition sighted *S*s, although still used less often than by Three-Comparison condition blind *S*s, U (7, 8) = 12.0, $p < .05$. At the same time, pinch usage by these sighted *S*s was similar to that of Three-Comparison blind *S*s, U (7, 8) = 23.5, $p > .10$. In the Five-Comparison condition, no significant scanning differences were indicated, a result perhaps accounting for the diminished error difference between blind and sighted *S*s in this condition compared to the Three-Comparison condition. (Top sweeping by all groups was low with no consistent changes as a function of Comparison-Array condition).

We summarized the percentage of judgments within successive four-trial blocks in which each strategy occurred to get a better view of the changes that took place in sighted *S*s' scanning as comparison array size increased. The variability of these scores made any meaningful statistical analysis difficult. However, the data for sighted *S*s did suggest a decrease in gripping over trials and an increase in pinching over trials in the One-Comparison condition, trends which did not occur in either the Three- or Five-Comparison conditions. There were no consistent effects of trials on top sweeping by sighted *S*s, or upon usage of any of the three strategies by blind *S*s.

Discussion

The present study extends Davidson's (1972a) paper in several ways. First, it appears that the contribution to perceptual accuracy made by exploratory search may be based upon an interaction between the task requirements and the style of search. This conclusion was suggested by Davidson (1972a) with respect to the nature of the stimulus attribute being judged, and seems extended by our data to also include the nature of the task paradigm.

Second, a scanning strategy that permits simultaneous prehension of a stimulus attribute seems to withstand increasing task difficulty better than methods that sample more successively (at least with regard to relational attributes such as curvature). Davidson (1972a, 1972b) has speculated that simulta-

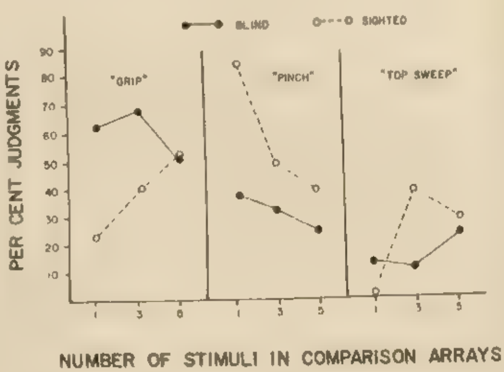


Figure 1. The percentage of judgments in which blind and sighted *S*s in Part I used the scanning strategies. (More than one strategy could occur in each of an *S*'s 24 trials. Mann-Whitney *U* tests revealed that blind *S*s gripped more than the sighted in the One-Comparison condition, $p < .05$, and Three-Comparison condition, $p < .05$, and the sighted pinched more than the blind in the One-Comparison condition, $p < .01$.)

neous prehension may help to pattern the relationship, reducing the requirement for integrating successive inputs in memory. If the increases in task difficulty in the present study are viewed as increases in task immediate recognition memory demand (as has been argued by Goodnow, 1971), then our results suggest that such scanning may also contribute to better recognition memory for form.

The greater accuracy of blind *S*s compared to the sighted is consistent with other studies of curvature categorization by the blind (Davidson, 1972a; Hunter, 1954), extending these previous results to the equivalence matching situation. This finding suggests that curvature is one form property that requires some haptic perceptual experience before it can be accurately perceived.

The link between perceptual accuracy and scanning style, found initially by Davidson (1972a) and replicated here, is a straightforward extension of previous findings for judgments of other form properties involving comparisons either of blind and sighted humans (Davidson, Barnes, & Mullen, in press; Hatwell, 1959) or young children (Abravanel, 1968). Moreover, the scanning differences between sighted *S*s depending on comparison array size suggests that modification of exploratory activity may be one result of perceptual experience.

REFERENCES

Abravanel, E. The development of intersensory patterning with regard to selected spatial dimensions. *Monographs of the Society for Research in Child Development*, 1968, **33** (2, Serial No. 118).

Davidson, P. Haptic judgments of curvature by blind and sighted humans. *Journal of Experimental Psychology*, 1972, **93**, 43–55. (a)

Davidson, P. The role of exploratory activity in haptic perception: Some issues, data and hypotheses. *Research Bulletin, American Foundation for the Blind*, 1972, **24**, 21–27. (b)

Davidson, P., Barnes, J., & Mullen, G. Differential effects of task memory demand on haptic matching of shape by blind and sighted humans. *Neuropsychologia*, in press.

Goodnow, J. Eye and hand: Differential memory and its effects on matching. *Neuropsychologia*, 1971, **9**, 89–95.

Hatwell, Y. Perception tactile des formes et organisation spatiale tactile. *Journal de Psychologie*, 1959, **56**, 187–204.

Hunter, I. Tactile-kinesthetic perception of straightness in blind and sighted humans. *Quarterly Journal of Experimental Psychology*, 1954, **6**, 149–154.

(Received July 26, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 691–695

PICTURE-WORD DIFFERENCES IN DISCRIMINATION LEARNING: I. APPARENT FREQUENCY MANIPULATIONS[1]

JOEL R. LEVIN[2]
University of Wisconsin—Madison

ELIZABETH S. GHATALA
Weber State College

LARRY WILDER
University of Wisconsin—Madison

A discrimination-learning task was administered to a sample of sixth graders. Stimulus materials consisted of either pictures or words for which frequency judgments had been obtained in a previous experiment. Picture and word items were selected which differed either substantially or minimally in apparent frequency measures. While the usual picture-over-word effect was found when picture–word apparent frequency differences were large, it vanished when pictures and words were equated in terms of apparent frequency. These results, coupled with some recent data, provide for a fairly complete account of picture–word differences in discrimination learning.

It has recently been demonstrated that pictures are more easily learned than words in a discrimination task, for both children (Rowe, 1972; Wilder & Levin, 1973) and adults (Rowe, 1972; Rowe & Paivio, 1971; Wilder & Levin, 1973). To date, one of the more popular explanations of this effect has been that pictures are more likely than words to produce a simultaneous verbal–imaginal encoding, and that dually coded materials are better recalled (Paivio, 1971). However, guided by the frequency theory of verbal discrimination learning originally proposed by Ekstrand, Wallace, and Underwood (1966), the present authors conducted an experiment which suggested that picture–word effects in discrimination learning may be attributable to differences in the subjective (apparent) frequency associated with the two types of material (Ghatala, Levin, & Wilder, 1973).

In that experiment, sixth graders were administered either a picture or word frequency judgment task in which line drawings of familiar objects or their printed verbal labels were presented for varying numbers of exposures, with *S*s later asked to estimate the number of times a particular picture or word had been previously shown. It was found that the mean judgments for pictures were higher than the mean judgments for words, even though the actual presentation frequencies of the two were the same. In addition, the average variability associated with items of the same presentation frequency was lower for pictures than words.

According to frequency theory, discrimination between the "correct" and "incorrect" members of a pair in a verbal discrimination task is based on the frequency differential between them, with more subjective frequency "units" accruing to the correct member of the pair than to the incorrect member through such factors as differential rehearsal favoring the former. The Ghatala et al. (1973) results are consistent with a frequency theory interpretation of picture–word differences

[1] This research was funded by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the National Institute of Education (Center No. NE-C-00-3-0015). The opinions herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred. The authors are grateful to the referees for improving the paper's clarity, to Becky Abbott for assistance in data collection, and to Cathy Bussey for typing the final draft of the paper. The order of the first two authors was determined randomly.

[2] Requests for reprints should be sent to Joel R. Levin, Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1025 West Johnson Street, Madison, Wisconsin 53706.

in discrimination learning since: (*a*) the finding that pictures were higher in mean apparent frequency than words suggests that a subjective frequency unit produced by a picture is larger than a frequency unit produced by a word; and (*b*) the finding that *S*s were less variable in estimating the presentation frequencies of pictures as compared to words suggests that differences in presentation frequency are more likely to be discriminated when the items are pictures than when they are words. Both findings lead to the prediction of superior discrimination learning when the materials are pictures—the first, because the absolute frequency difference between correct and incorrect items is larger for picture pairs than for word pairs; the second, because lower variability associated with correct and incorrect items in picture pairs renders them more discriminable than word pairs.

The present experiment was conducted to test the hypothesis that picture–word differences in discrimination learning are a function of apparent frequency differences associated with the two types of material. How apparent frequency judgments of pictures and words were obtained and manipulated is discussed in the following section.

Method

In the Ghatala et al. (1973) study, 19 sixth-grade *S*s judged the frequencies of 39 pictures, 20 of which had been presented once, 10 twice, 6 three times, and 3 four times. Another 19 *S*s gave frequency judgments for the printed labels which occurred with the same presentation frequencies as their corresponding pictures.

Manipulation of apparent frequency means. A comparison of the mean apparent frequencies of individual picture and word items revealed that the mean of pictures was reliably ($p < .001$) higher. However, while the mean judgments of pictures (at each presentation frequency) were, in general, higher than those of words, there was some overlap in the distributions. In particular, from among the items presented once, 8 of the pictures could be matched fairly closely with 8 of the words. Similarly, 4 pictures and 4 words were selected from among the items presented twice, and 2 pictures and 2 words were selected from among the items presented 3 times. Pairing items within presentation frequencies resulted in 7 picture pairs and 7 word pairs which differed minimally in mean apparent frequency (small-difference items). The mean apparent frequency of the pictures used in the small-difference items (averaged over the 14 pictures) was 1.42. The mean apparent frequency of the 14 words also averaged 1.42.

Next, 14 pictures and 14 words were selected to form pairs which differed considerably in mean apparent frequency (large-difference items). Four picture pairs and 4 word pairs were formed from items presented once, 2 pairs of each stimulus type were formed from items presented twice, and 1 pair apiece were formed from the items presented 3 times. The mean apparent frequency of the pictures was 1.74, and that of the words was 1.16.

The 14 picture pairs (7 small-difference items plus 7 large-difference items) comprised 1 discrimination list, and the 14 word pairs comprised another. In forming each list, an attempt was made to equate a pair's correct and incorrect members in apparent frequency. This was done in order to ensure that any between-materials (pictures and words) apparent frequency differences were manipulated independently of within-materials (correct and incorrect members) apparent frequency differences.

Three presentation orders of each list were constructed such that: (*a*) within each order, the occurrence of small- and large-difference items was random; (*b*) within an order, the correct member was located on the right for half of the pairs and on the left for the other half; and (*c*) across orders, the spatial location of the correct member of each pair was varied randomly with the restriction that it not occur in the same position across all 3 orders. The randomization scheme for constructing the 3 presentation orders was the same for picture and word lists, i.e., each order's sequence of small- and large-difference items, as well as its spatial location of correct members, was the same for the 2 lists.

The pictures were line drawings (3 × 3 in.) which were pasted side by side on sheets of 8½ × 11 in. paper. The sheets were then placed in a 3-ring notebook binder. The words were typed in primary type, side by side, on sheets of paper which were placed in a separate 3-ring binder. Asterisks were used to indicate the correct member in each pair.

Manipulation of apparent frequency variances. From the same pool of items used to construct lists varying in mean apparent frequency, picture and word items were selected on the basis of the variance in apparent frequency associated with each item. The variance of these judgments differed for pictures and words, being statistically smaller ($p < .001$) for pictures. Once again, there was some overlap between the distributions. Small-difference items were constructed from 14 pictures (8 selected from among the "one" items; 4 from among the "two" items, and 2 from among the "three" items) and 14 words (selected in the same numbers as the pictures from the 3 presentation frequencies) which differed little in variability. Within presentation frequencies, picture pairs and word pairs were formed which differed minimally

in variance. The average variance of the pictures was .54, and that of the words was .52.

Fourteen pictures and 14 words which differed widely in variance (lower for pictures) were used to construct the large-difference items. The pictures and words were selected from presentation frequency levels in the same numbers given above. As in all previous cases, pairs were constructed from items within a presentation frequency. The average variance for the 7 picture pairs so constructed was .39. The mean variance for the word pairs was .80.

As before, the 7 small-difference and 7 large-difference items within each stimulus type were combined to form a 14-pair picture list and a 14-pair word list. The procedures used in creating the lists were the same as those previously described.

Finally, it should be noted that in selecting items to investigate one of the measures of apparent frequency (either means or variances) an attempt was made to control the other measure, even though this could not be done perfectly. For example, the average variances of picture and word items selected for a small difference in mean judgments were .45 and .65, respectively. For large-difference items, these figures were .55 and .61. With regard to items selected for small differences in variance, the mean apparent frequencies were 1.64 for picture pairs and 1.40 for word pairs. For the items with large variance differences, the corresponding means were 1.49 and 1.41.

Subjects. The *Ss* were 80 sixth-grade children (mean age = 12.4 yr.) randomly selected from two elementary schools located in middle-class neighborhoods in Ogden, Utah. Twenty *Ss* were assigned (according to a block-randomized procedure) to each of the 4 conditions formed by the combination of 2 materials (pictures or words) and 2 measures (means or variances). Following Walster and Cleary's (1970) suggestions for equating "statistical" significance with "practical" significance, the number of *Ss* required was based directly on the simultaneous specification of effects considered important (differences in excess of 1 within-cells standard deviation for the large-difference items) and those considered trivial (differences less than $\frac{1}{4}$ of a within-cells standard deviation for the small-difference items).

Results

Performance on the discrimination task is summarized in Table 1. For both mean and variance manipulations, the results are clear-cut. With large-difference items, pictures and words differ substantially: means, $F(1, 38) = 13.58$, $p < .001$; variances, $F(1, 38) = 6.48$, $p < .05$; whereas with small-difference items, they do not ($F < 1$ for both means and variances).

TABLE 1

Mean Number of Correct Discriminations in Two Trials on Items with Large or Small Mean and Variance Differences (Maximum Score = 14)

Measure	Large difference		Small difference	
	Pictures	Words	Pictures	Words
Means	12.30	10.20	11.70	11.45
Variances	12.15	10.45	11.80	11.60

While there was marked improvement over trials, in no case did this interact with the above picture–word effects (all ps $> .05$).

Discussion

Based on the assumption that discrimination learning consists of discriminations of subjective frequency differences between items (Ekstrand et al., 1966), and given the empirical finding that apparent frequencies associated with pictures and words differ (Ghatala et al., 1973), we were able to anticipate the usual picture-over-word superiority in discrimination learning (Rowe, 1972; Rowe & Paivio, 1971; Wilder & Levin, 1973) when apparent frequency differences between pictures and words were maintained. On the other hand, when pictures and words were equated in terms of either apparent frequency means or variances (as determined from an earlier study), no discrimination-learning differences between the two types of material were detected.

When one considers the (almost incredibly) straightforward pattern of results, it is tempting to pit plausible rival constructs against the authors' preferred account of the effects in terms of apparent frequency differences. Thus, it might be argued that when apparent frequencies were manipulated to create large- and small-difference items, other attributes were inadvertently manipulated concomitantly. Normative word frequency (f) as indexed by the Thorndike–Lorge (1944) norms, and measures of meaningfulness (m) and concreteness (C)—all of which have been found to influence the ease of verbal discrimination learning—may have been differentially represented in the large- and small-difference word pairs. However, based on an inspection of the present words for which norms were available (Paivio, Yuille, & Madigan, 1968), it was found that all the words were uniformly

high in f, m, and C, with no discrepancies between large- and small-difference items evident. Of course this finding was to be expected given that the present stimuli were selected to be picturable (i.e., concrete) and familiar to elementary school children (i.e., frequent and meaningful).

While there are potentially a multitude of uncontrolled factors which may have contributed to the observed effects, a reviewer of this journal has pointed out one in particular which will be considered in detail. If it is assumed that apparent frequencies of pictures are correlated with those of their verbal labels (the corresponding words), *and* if it is further assumed that the words corresponding to, say, large-difference pictures are not equally divided between large- and small-difference items (due to nonrandom assignment of items to conditions), then it is possible in the present study that *apparent frequency* has been confounded with the *specific items* appearing in the various conditions. For example, if the large-difference pictures (which tend to produce higher and less variable judgments than the small-difference pictures) correspond to words with high means and low variances, *and* if more of these words happened to be assigned to the small-difference than to the large-difference condition, then the pattern of results in Table 1 may be easily explained in terms of item-specific peculiarities; that is, large-difference pictures would be expected to be best, large-difference words worst, with the two other conditions in between. The same argument can be made if words corresponding to small-difference pictures ended up more frequently in the large-difference than in the small-difference condition.

Fortunately, a thorough examination of the present materials suggests that this account of the results is not satisfactory. In the first place, for the items equated on the variance measure, the words corresponding to large-difference pictures were almost equally divided between the large- and small-difference conditions, as were the words corresponding to small-difference pictures (a total of 10 vs. 12 cases, in the direction *opposite* to the hypothesis). Second, although an inspection of the items equated on means indicated a tendency toward unequal division in the direction specified by the hypothesis (a total of 15 vs. 6 cases, $p < .10$), further scrutiny of the Ghatala et al. (1973) data revealed that the correlation between picture and word items presented once was low ($r = .14$) and statistically nonsignificant, thereby mitigating against the reviewer's first assumption. Thus, the specific-item hypothesis, with its assumptions of picture–word correlations and selective item assignment, cannot account for the pattern of results observed here: In the case of the variances, there was no differential item selection; and in the case of the means, the picture–word correlation was low.

While it is not possible to state with certainty which of the two measures, means or variances, has the more predominant influence on picture–word differences in discrimination learning (since the two measures covary, and this problem was not completely controlled for here), speculations may be made on the basis of some previous research. In one published study (Ghatala & Levin, in press-a) and in other unpublished work, it has been found that pictures and words differ more reliably with respect to variability in apparent frequency (pictures being lower) than with respect to mean apparent frequency. If this result continues to be replicated in further controlled experiments, it would indicate that pictures are easier to discriminate than words because the subjective frequency units associated with them are more stable.

Even though a frequency theory (Ekstrand et al., 1966) explanation of the present effects is appealing—in particular, with regard to the operationalization of the manner and contexts in which frequency units accrue—it must be realized that the validity of an apparent-frequency construct is not enhanced unless one can explain *why* pictures and words differ in apparent frequency. However, such an attempt has been made by Ghatala et al. (1973). Essentially, the argument is that pictures and words differ substantially in terms of *S*'s preexperimental encounters with them. Thus, while *S*s have likely read the experimental words on numerous preexperimental occasions, they have never encountered (preexperimentally) the particular experimental line drawings. By adapting Weber's law to this type of situation, it is readily understandable why the presentation of a picture (initially low in preexperimental exposure) should constitute a larger—and, evidently, more stable—subjective frequency unit than the presentation of a word (initially high in preexperimental exposure).

Based on this rationale, some interesting extensions have been made (cf. Ghatala & Levin, in press-a, in press-b). However, there are additional verbal discrimination and recognition memory phenomena with which the

"background-frequency" explanation must deal before it can be considered viable. Whether or not it can negotiate them successfully remains to be seen.

REFERENCES

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **73**, 556–578.

Ghatala, E. S., & Levin, J. R. Developmental differences in frequency judgments of words and pictures. *Journal of Experimental Child Psychology*, in press. (a)

Ghatala, E. S., & Levin, J. R. Discrimination learning as a function of differences in materials: A proposed explanation. *Memory and Cognition*, in press. (b)

Ghatala, E. S., Levin, J. R., & Wilder, L. Apparent frequency of words and pictures as a function of pronunciation and imagery. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 85–90.

Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.

Paivio, A., Yuille, J. C., & Madigan, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Pt. 2).

Rowe, E. J. Discrimination learning of pictures and words: A replication of picture superiority. *Journal of Experimental Child Psychology*, 1972, **14**, 303–312.

Rowe, E. J., & Paivio, A. Discrimination learning of pictures and words. *Psychonomic Science*, 1971, **22**, 87–88.

Thorndike, E. L., & Lorge, I. *The teacher's wordbook of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

Walster, G. W., & Cleary, T. A. Statistical significance as a decision-making rule. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), *Sociological methodology.* San Francisco: Jossey-Bass, 1970.

Wilder, L., & Levin, J. R. A developmental study of pronouncing responses in the discrimination learning of words and pictures. *Journal of Experimental Child Psychology*, 1973, **15**, 278–286.

(Received June 26, 1973)

— — —

(Continued from page 686)

Category Similarity and Retroactive Inhibition in Free Recall. Isabel M. Birnbaum (Department of Psychology, University of California, Irvine, California 92664).

Interaction of Positive and Negative Labels with Category Composition in Attribute Identification Concept. Peder J. Johnson (Department of Psychology, University of New Mexico, Albuquerque, New Mexico 87131) and Thomas C. Toppino.

The Effect of Delayed CS Termination on Extinction of an Avoidance Response following Different Termination Conditions. Allen C. Israel (Psychology Department, State University of New York at Albany, 1400 Washington Avenue, Albany, New York 12222), Vernon T. Devine, Margaret A. O'Dea, and Mark E. Hamdi.

Assessment of Selective Search as an Explanation for Intentional Forgetting. Donald Homa (Division of Natural Sciences, New College, Sarasota, Florida 33578) and Susan Spieker.

Adaptation Level and the Central Tendency Effect in Stimulus Generalization. David R. Thomas (Department of Psychology, University of Colorado at Boulder, Muenzinger Building, Boulder, Colorado 80302), Harry Strub, and James F. Dickson, Jr.

The Spacing Effect and the A–B, A–C Paradigm: Evidence for Retroactive Facilitation. Donald Robbins (Department of Psychology, Emory University, Atlanta, Georgia 30322) and James F. Bray.

Class Structure in the Biasing of Perceived Pattern Similarity. Leona S. Aiken (Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122), Richard M. Fenker, and Selby H. Evans.

Compensatory Effects in Moral Judgment: Two Rights Don't Make Up for a Wrong. Dwight R. Riskey (Department of Psychology, University of California, Los Angeles, California 90024) and Michael H. Birnbaum.

Simultaneous Discrimination: Two Separate Types of Control. Michael D. Zeiler (Department of Psychology, Emory University, Atlanta, Georgia 30322).

Facilitative and Interference Effects of Response Grouping in Paired-Associate Learning. Peggy A. Runquist (Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E1, Canada) and Gerrit O. Aronson.

Pronunciation Effects in Verbal Discrimination Learning. Larry Wilder (Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1025 West Johnson Street, Madison, Wisconsin 53706), Joel R. Levin, Michael Kuskowski, and Elizabeth S. Ghatala.

Syntagmatic versus Paradigmatic Paired-Associate Acquisition. Larry Riley (Department of Psychology, Northwest Missouri State University, Maryville, Missouri 64468) and Gary Fite.

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 696–699

A PREVIEW-CONSTRAINT MODEL OF ROTARY ARM CONTROL AS AN EXTENSION OF FITTS'S LAW [1]

TARALD O. KVÅLSETH [2]

School of Industrial and Systems Engineering, Georgia Institute of Technology

A first-order linear model relating movement time to Fitts's movement variable and a preview variable was formulated and validated experimentally using 15 undergraduate students as *S*s. Without preview constraint, the rotary arm movement task produced a maximum movement information or index of task difficulty (ID) rate of 4.7 bits/sec, which, when preview constraints were imposed, was reduced to 3.9 bits/sec as compared to a marginal preview ID rate of 12.5. The movement variable was found to account for about 70% of the total contribution to movement time. The error rates were determined to be highest in the no-preview-constraint case and were significantly ($p < .05$) affected by both movement and preview ID.

The interrelation between movement time (Y), amplitude (A), and accuracy or target width (W) for terminating the movement for serial motor tasks was first formulated within an information-theoretic context by Fitts (1954). Arguing by analogy from Shannon's channel capacity Theorem 17 (Shannon & Weaver, 1949), Fitts proposed the simple relationship

$$Y = a + bX_1 + \epsilon; \qquad [1]$$
$$X_1 = \log_2(2A/W) \text{ bits},$$

where X_1 is considered to be a measure of the amount of entropy or information in each movement, ϵ is the error term, and a and b are constant parameters. The validity of this linear model has been clearly confirmed by Fitts and other investigators (cf. Welford, 1968, Chap. 5) through a number of experiments such as those involving tasks for which *S*s were required to move a stylus or pen quickly between 2 targets of width W and distance apart A.

Fitts's model was concerned with linear arm movements without any constraint on the movement path or the preview. A linear second-order model for hand movements with path constraints was recently formulated by Kvålseth (1973); however, as far as this author is aware, no quantitative model has so far been formulated and validated that incorporate preview constraints for serial motor movements involving, specifically, arm rotations (Crossman, 1960, studied the effect of preview on continuous tasks involving pursuit tracking). The present investigation was designed to study the effect of such variable preview in laboratory experiments.

The basic hypothesis to be tested was that the movement time Y is linearly related to a new index of task difficulty (ID) variable

$$X = \log_2\left[\gamma\left(\frac{2A}{W}\right)^{\alpha}\left(\frac{1}{P}\right)^{\beta}\right], \qquad [2]$$

where P is the preview (distance) and α, β, and γ are unknown constant parameters. This model may be placed within an information-theoretic context that represents essentially a generalized extension of Fitts's interpretation of the variable X_1 (cf. Fitts & Posner, 1967, pp. 112–117) without proposing that it be considered as a necessarily true paradigm of human behavior relevant to the motor task considered in this study. Thus, the quantity X may be interpreted as the information generated by or the entropy associated with the selection of any one movement from an ensemble of equiprobable movements; the term in the square bracket of Equation 2 being the number of elements in the en-

---

[1] The author wishes to thank W. B. Cross for assisting with data collection.

[2] Requests for reprints should be sent to Tarald O. Kvålseth, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332.

semble. The unit of X may then be considered as being "bits" with the constraint imposed that all values of the different variables and parameters considered are such that $X > 0$.

Effectively, the variable P is such that $0 < P \leq (A + W/2)$. For the case of no preview constraint, $P = (A + W/2)$ or alternatively, it may be considered that $\beta = 0$, so that the resulting expression for X is simply a generalization of X_1 in Equation 1; the identical linear model of Equation 1 is still seen to apply. For the preview-constraint case, the linear relationship between Y and X can, after a few simple algebraic manipulations, be expressed alternatively in terms of the following first-order linear model,

$$Y = a + bX_1 + cX_2 + \epsilon; \qquad X_2 = \log_2(1/P) \tag{3}$$

where the ID variable X_2, which may assume negative values and did so exclusively in the present study, is not to be interpreted as a direct information or entropy measure in the information-theory sense. The coefficients a, b, and c are unknown constant parameters to be estimated, and ϵ is the experimental error.

Method

Apparatus and task. The experimental apparatus designed for this experiment is illustrated in Figure 1. It consisted of a black box containing a 1½-v. battery that provided the input voltage to a linear potentiometer whose output terminals were connected to a single-channel strip chart recorder. The potentiometer was fixed to the inside of the box with its rotary wiper or shaft emerging through a hole in the center of one side of the box. This shaft was attached to a round control knob 2¼ in. in diameter so that any angular displacement of the control knob produced a proportional variation in the output voltage of the potentiometer, which was thus used as a displacement transducer. A 2½-in.-long pointer (measured from the center of the control knob and the shaft) was mounted onto the knob such that it barely cleared a scale attached to the box. Thus, a control–display ratio of 1 was used. Such an easily replaceable scale, which was placed such that the end of the black pointer reached approximately halfway across it throughout its length, had a marked target area of width W degrees; the centers of the target areas were separated by a distance A degrees.

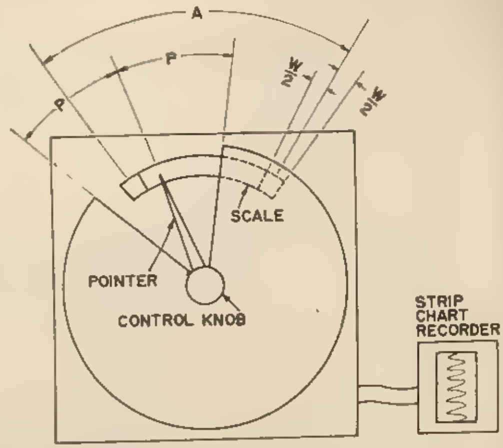


FIGURE 1. A plan view of the apparatus.

Preview constraint was incorporated into the experiment by attaching a black circular cardboard plate to the bottom of the control knob. The diameter of the plate was slightly larger than that of the outer side of the scale measured from the center of the control knob (see Figure 1). This plate, which was easily removable, had a radial slit in it of $2P$ degrees and was placed such that the pointer was in the center of the slit. Several such plates and scales were designed to provide different experimental conditions involving different values of A, W, and P.

The apparatus was placed on the top of a table, and *Ss* stood in an upright position. The experimental task required *S* to move the pointer back and forth between the 2 target areas of a scale by turning the control knob; this movement had negligible friction. Because of the preview constraint, *S* could only see the scale a distance of P degrees ahead of the pointer throughout a complete movement between targets. For such serial movements, which were simultaneously recorded on the strip chart, *Ss* were told to minimize their movement times, while at the same time trying to maintain low error rates. An error was considered as being committed whenever a pointer movement terminated outside 1 of the 2 targets.

Subjects. A total of 15 unpaid *Ss* were used in the experiment, all of whom were right-handed male undergraduate students. Five of these performed the experiment without any preview constraint, while the remaining 10 *Ss* were subjected to preview constraints.

Experimental procedure. The 12 experimental conditions used consisted of all combinations of angular movement amplitudes $A = 30°, 60°, 90°$ and target widths $W = 5°, 10°, 15°, 20°$, with preview $P = W$. The order of these conditions was randomized and different for each *S*. Prior to each run, the pointer was lined up with each of the 4 boundary lines of the target pair, and the strip chart was run a short distance for each of these 4 pointer settings so as to identify the equivalent target

boundaries on the strip chart. Then, with the strip-chart recorder turned off, S was given a 10-sec. practice session. Following a practice session and an approximate 5-sec. rest period, S performed the experiment for 15 sec. while the strip-chart recorder was turned on. From this chart, E was able to obtain the total number of moves made by S and hence the mean movement time during the 15-sec. interval. The chart also provided information about the error rate by simply observing the fraction of moves by the recorder pen that fell short or far of the target boundaries. Movements that resulted in errors were included in the determination of the mean movement time.

Results and Discussion

No-preview-constraint case. On the basis of least squares regression, the relationship between rotary arm movement time and movement ID was estimated as

$$\hat{Y} = -21.02 + 21.31\,X_1, \qquad [4]$$

where the unit of $\hat{Y}$ is 10^{-2} sec. The parameter estimates are based on average movement time over each 15-sec. period and then over Ss. The coefficient of multiple determination for this relationship was determined to be $R^2 = .96$, indicating an excellent fit for this regression or a highly significant ($p < .0005$) correlation between Y and X_1 of $r = .98$ (summary data are given in Table 1). On the basis of these data, the maximum information rate for this type of movement task is seen to be constant for different X_1 values at approximately 4.7 bits/sec; $1.58 \leq X_1 \leq 5.17$ in this experiment. This information rate is slightly lower than that for linear serial hand movements (determined by Kvålseth, 1973, to be 5.5 bits/sec) and .4 times that for arm movements in reciprocal tapping tasks (cf. Fitts, 1954).

Preview-constraint case. Least squares estimation for Equation 3 based on average movement time in 10^{-2} sec. (averaged for each S during 15-sec. period and then averaged across the 10 Ss) resulted in the following relationship,

$$\hat{Y} = 8.97 + 26.06\,X_1 + 8.03\,X_2. \qquad [5]$$

This first-order model gives an excellent fit to the experimental data with $R^2 = .98$. The regressor X_1 is by far the most "significant" of the 2 regressors in terms of its contribution to (the calculated valu[illegible] Specifically, on the basis of the [illegible] determination coefficients, which [illegible] found to be .93 and .41 for X_1 and [illegible] respectively, X_1 accounts for appr[illegible]mately 70% of the total contribution to [illegible] by the 2 regressors. However, as is se[illegible] from the data in Table 1, the addition[illegible] effect of X_2 is significant at the .05 [illegible]

From Equation 5, it seems plausib[illegible] to test the statistical hypothesis that $c = [illegible]/3$. The appropriate t statistic for this test was found to take on the value .20 with 9 degrees of freedom so that this hypothesis is clearly accepted at the .05 level. Thus, Equation 3 may be replaced by the model

$$Y = \beta_0 + \beta_1 X_3 + \epsilon;$$
$$X_3 = \log_2(2A/WP^{\frac{1}{3}}), \qquad [6]$$

where β_0 and β_1 are unknown constant parameters and ϵ denotes the error term. The ID variable X_3 is then the equivalent of X_1 when preview constraint is imposed and, although it is seen that X_3 may possibly assume negative values, only positive X_3 values were used in the present study.

On the basis of a least squares regression of Y on X_3, employing the same values of the variables Y, A, W, and P as those used in arriving at Equation 5, the estimates for β_0 and β_1 were determined to be

$$\hat{\beta}_0 = 12.00, \quad \hat{\beta}_1 = 25.72,$$

providing an R^2 of .97; the analysis of variance data corresponding to the test of the hypothesis that $\beta_1 = 0$ are given in Table 1. The resulting rate $dX_3/d\hat{Y} = 3.9 \approx \partial X_1/\partial\hat{Y}$ is seen to be 17% lower than the maximum information rate for the no-preview-constraint case. Furthermore, the marginal ID rate $\partial X_2/\partial\hat{Y} = 12.5$ is observed to be comparable to the maximum ID rate obtained by Fitts (1954) for linear arm movements.

Error rates. The percentage of movements that missed the indicated targets tended to be consistently higher in the no-preview-constraint case than when preview constraints were imposed, although the instructions to Ss were the same in the 2 cases. In the former case, the

average error rate (average across Ss for each of the 12 experimental conditions) ranged approximately 5%–15% and was highest for large X_1 values without appearing to be entirely systematically related to X_1. The average of this error rate across the X_1 values used was found to be about 10%, which was twice that of the preview-constraint case. In general, the errors appeared to be about equally distributed between overshoot and undershoot errors. For the preview-constraint case, it was established that task difficulty, as reflected by values of X_1 and X_2, had a significant effect on the error rate. The average error rates for each S were paired according to low index of difficulty level ($1.58 \leq X_1 < 3.58$ and $-4.32 \leq X_2 < -3.32$) or, equivalently ($.14 \leq X_3 < 2.48$) and high difficulty level ($3.58 \leq X_1 \leq 5.17$ and $-3.32 \leq X_2 \leq -2.32$) or, equivalently ($2.48 \leq X_3 \leq 4.40$) and, on the basis of the Wilcoxon signed-ranks test, the hypothesis of no difference between these levels was rejected at the .05 level.

TABLE 1

Analysis of Variance

Source	df	F
No-preview-constraint case (Equation 4)		
X_1	1	237.43**
Residual (*MS*)	10	(22.78)
Total	11	
Preview-constraint case (Equation 5)		
X_1	1	243.06**
Addition of X_2	1	6.21*
X_1, X_2	2	
Residual (*MS*)	9	(30.24)
Total	11	
Preview-constraint case (Equation 6)		
X_3	1	416.82**
Residual (*MS*)	10	(27.27)
Total	11	

* $p < .05$.
** $p < .001$.

Conclusion

This study has generalized Fitts's law for simple linear and serial arm movements to incorporate temporal motor responses involving rotary arm movements with preview constraints. It was demonstrated that the imposition of preview constraints on such peripheral motor tasks had a significant effect on movement time, error rates, and maximum information rate. The simple first-order linear models relating movement time to movement variable X_1, the preview variable X_2, and the composite variable X_3 were clearly appropriate formulations providing excellent fits to the experimental data generated. Of the 2 variables, X_1 and X_2, the latter was found to be the least "important" one in explaining the variation in movement time; its contribution to movement time was determined to be less than half that of X_1. Further observations resulting from imposing preview constraints were that the average error rate was reduced by some 50% while the maximum information rate was 17% below that of the no-preview-constraint case.

No attempt was made in the present study to determine if practice had any influence on the degree to which the preview constraint effected movement time. It is conceivable that there is an interaction between practice and preview so that the preview constraint has a greater effect on movement time early during the task performance than it has later when S may not have to rely as much on preview information. This point ought to be worth pursuing in another study.

The effect of preview constraints on the task parameters as determined in this study applies to simple, serial, and 1-dimensional rotary arm movements and may be quite different for other types of motor tasks. A further investigation is planned with objectives similar to the present study but involving linear movement tasks with and without different types of movement path constraints.

REFERENCES

Crossman, E. R. F. W. The information-capacity of the human motor-system in pursuit tracking. *Quarterly Journal of Experimental Psychology*, 1960, **12**, 1–16.

Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 1954, **47**, 381–391.

Fitts, P. M., & Posner, M. I. *Human performance.* Belmont, Calif.: Brooks/Cole, 1967.

Kvålseth, T. O. Fitts' law for manipulative temporal motor responses with and without path constraints. *Perceptual and Motor Skills*, 1973, **37**, 427–431.

Shannon, C. E., & Weaver, W. *The mathematical theory of communication.* Urbana: University of Illinois Press, 1949.

Welford, A. T. *Fundamentals of skill.* London: Methuen, 1968.

(Received July 30, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 700–705

HABITUATION OF THE DIGITAL VASOCONSTRICTIVE ORIENTING RESPONSE

STEN E. LEVANDER[1]
Department of Psychiatry, Karolinska sjukhuset, Karolinska Institutet, Stockholm, Sweden

LARS LIDBERG
Laboratory for Clinical Stress Research, Karolinska sjukhuset, Stockholm, Sweden

DAISY SCHALLING[2]
Department of Forensic Psychiatry, Karolinska Institutet, Stockholm, Sweden

Changes in finger volume (FV) and finger pulse volume (PV) were measured by two pneumoplethysmographic methods. Eleven identical randomly spaced auditory stimuli were given. Five different habituation estimates were calculated; H-score, parametric beta, parametric correlation, Kendall's tau, and nonparametric beta. Both FV and PV responses habituated, PV responses faster than FV. The correlations between FV and PV responses were significant and moderately high. Some surprisingly low correlations were obtained among the different habituation estimates, indicating that the choice of habituation estimate is of great importance, and might influence the result profoundly. Differences in response latency found between FV and PV responses were discussed in terms of measurement technique.

Much psychophysiological research in recent years has been directed toward exploring empirical and theoretical aspects of habituation of orienting responses in different autonomic systems (Frankenhaeuser, Fröberg, Hagdahl, Rissler, Björkvall, & Wolff, 1967; Groves & Thompson, 1970; Lader & Wing, 1966; McDonald, Johnson, & Hord, 1964). Speed of habituation of electrodermal responses has been found to be a highly stable individual characteristic (Crider & Lunn, 1971) and there is evidence for at least moderate genetic determination (Lader & Wing, 1966).

Habituation is in general rapid for skin conductance (SC) responses. In regard to habituation of the digital vasoconstrictive response the evidence is equivocal. Davis, Buchwald, and Frankmann (1955) found significant habituation for pulse volume (PV) but not for finger volume (FV) responses. This dissociation was interpreted to indicate that PV and FV changes are dependent on different physiological systems. The difference in response latency found between the two measures was regarded by these authors as further evidence for this hypothesis. Their conclusions have been cited in subsequent review, e.g., Martin (1961). Burch (1961) also found habituation for PV responses. Unger (1964) found that 12 of 20 subjects reached a habituation criterion of three successive PV nonresponses to successive number of stimuli up to 36 stimulus presentations. Using the same habituation criterion Koepke and Pribram (1967) found that about half of the subjects habituated in PV, also with a variable number of tone stimuli up to 40. However, in a study by Lader (1965) there was no significant difference between the first four and the last four PV responses to repeated tone stimuli. Hare (1968) found indications of great intraindividual variability and resistance to habituation for PV responses in criminal subjects. Further, he noted that it was not unusual for several nonresponses to be followed by a number of responses of appreciable size.

The purpose of the present study was to investigate simultaneously the habituation of FV and PV responses to repeated auditory stimuli by a direct pneumoplethysmographic technique. Since there is a tendency to dishabituation in the vasoconstrictive orienting response, the use

[1] Requests for reprints should be sent to Sten E. Levander, Department of Psychiatry, Karolinska sjukhuset, S-104 01 Stockholm 60, Sweden.
[2] Also at the University of Stockholm, Sweden.

of the simple habituation criterion of three successive nonresponses is questionable. Therefore different methods of estimating habituation were applied.

Method

Procedure

Subjects. Nineteen conscripts, 19–21 yr. old, served as *Ss*. Participation was voluntary and *Ss* were paid. The *Ss* were in good physical and mental health.

Auditory stimuli. Eleven sine-wave tones of 1,000 Hz., approximately 100-db. intensity, and 1-sec. duration were administered through a loudspeaker. This was placed 1 m. to the right of the subject. The intervals between the tones were randomized and varied between 35 and 60 sec.

Measurement of finger volume. A pressure recording system (Elema EMT 32) with a pressure transducer of capacitor microphone type and amplifier with a stepwise variable time constant was used. The transducer was connected to a plastic oncometer cuff via a short rigid plastic tube. The cuff was mounted airtight on the left long finger of the subject. To eliminate the fast pulse volume deflections, the recordings were made with a high time constant (3 db. down at .3 Hz.).

Measurement of pulse volume. A pressure transducer of piezoelectric type (Elema EMT 510, see Lund, 1964) was connected to a plastic oncometer cuff via a short rigid plastic tube. The cuff was mounted airtight on the left index finger of the subject. Time constant for the outfit was 2 sec. (transducer connected directly to the galvanometer amplifier of the Mingograph).

Recording instrument. An inkwriter of the Elmqvist fast-response type (Mingograph 81) was used. As a control variable the left-hand long-finger temperature was continuously monitored by a thermistor device.

Procedural detail. The experiment was performed in September, a period with stable weather conditions. The barometric pressure varied between 747- and 763-mm. Hg. The *Ss* came to the laboratory in the afternoon, after their normal duties. They were placed in an examination chair in a quiet room, in which the temperature varied between 21.5° and 23.5° C. The finger temperature of the subjects varied between 21.5° and 34.8° C. at the beginning of the stimulation period. A constant, faint, indirect light was used. The recording apparatus was kept in an adjoining room, from which the experimenter could observe the subject through a small window.

The left arm and hand of the subject was placed on a soft support so that the limb was in a position of rest, and so that the fingers were at the level of the sternal angle.

The procedure was explained to the subject, and several deep inspirations were practiced. Then the recordings were started. Thereafter all instructions to the subject were given through a tape recorder via a loudspeaker.

The experiment started with 10 min. of rest. Then the 11 tone stimuli were administered.

Treatment of Data

Of the measures obtained, only data pertaining to habituation of responses are reported in this paper (cf. Lidberg, Schalling, & Levander, 1972).

Finger volume. Any observable decrease of the finger volume occurring between 1.6 and 10 sec. from the onset of the stimulus was considered as a response. Responses within 1.6 sec. were disregarded. (There were eight such responses, 5% of the total number of responses.)

A response was quantified as the difference, in mm. recording paper, between the maximal deflection during 10 sec. after the stimulus and the mean level during 10 sec. before the stimulus. This value was then corrected for transducer sensitivity, electrical amplification, and the volume of the finger tip inserted into the cuff.

Pulse volume. Any observable decrease of the pulse volume occurring between 1.6 and 10 sec. from the onset of the stimulus was considered as a response. Responses within 1.6 sec. were disregarded. (There were six such responses, 4% of the total number of responses.)

A response was quantified as the difference, in mm. recording paper, between the mean of the three smallest consecutive deflections during 10 sec. before the stimulus, and the mean of the three smallest deflections during 10 sec. after the stimulus. The difference was expressed in percent of the mean of the three smallest deflections during 10 sec. before the stimulus.

Response latency. As the time constant of the two recordings differed, no meaningful comparisons between response latencies of FV and PV were possible.

Habituation. Five estimates of habituation were calculated. Parametric correlations and regressions were calculated for response size on the logarithm of stimulus numbers. The third of a series of 3 successive nonresponses were regarded as the end of a response sequence. Most subjects, however, did not reach that criterion, and thus their response sequence consists of 11 responses. As there was a significant positive correlation between slope and intercept (estimated first response) for the parametric regression equations, a habituation estimate (H score) according to the procedure described by Montagu (1963) was calculated, in which correction for differences in estimated first response is applied.

As the use of parametric statistics requires normal distributions, homoscedasticity, and linearity of regression equations, and as these requirements are not strictly satisfied in the present data, as in most other studies, two nonparametric methods were also applied. Kendall's tau is a nonparametric correlation coefficient, which is relatively insensi-

tive to ties as compared with the Spearman rank correlation coefficient. However, a correlation coefficient estimates not only the slope of an assumed regression curve, but also deviations of responses from the curve. Sen (1968) has described a nonparametric regression statistic, nonparametric beta, based on the theoretical work of Kendall. Calculation of individual confidence limits is possible.

Thus five main habituation estimates were calculated on the present data, three parametric estimates based on the plot of response size on log stimulus number—product-moment correlation (r), linear regression (beta and H score)— and two nonparametric estimates—Kendall's tau and nonparametric beta. Further, some other simple estimates of habituation were calculated to allow for a direct comparison with the findings in some earlier studies (t test between means of first four and last four responses, the Wilcoxon test, and number of responses to the criterion of three successive nonresponses).

Results

Mean responses to tones in FV and PV, expressed as percent of the first response of each subject are shown in Figure 1. The trend appears to be linear and decreasing for both measures, though it appears more pronounced for PV than for FV. The slopes of the regression lines were −.40 for FV and −.55 for PV.

The correlations between mean responses in FV and PV were .62 over subjects and .44 over stimuli. For individual subjects the correlations varied between −.22 and .98, 25% of the subjects having correlations above .90. Two subjects who had many spontaneous fluctuations obtained very low correlations, −.22 and .11.

The distributions of H score, parametric beta, tau, and nonparametric beta justified the use of the t test to estimate if there was a significant habituation trend for the group in FV and PV. The t values for mean habituation are seen in Table 1. For PV, all t values were highly significant, whereas for FV two estimates (tau and H score) were significant, one was almost significant (parametric beta), and one was nonsignificant (nonparametric beta). For FV, no subject reached the habituation criterion of three successive nonresponses.

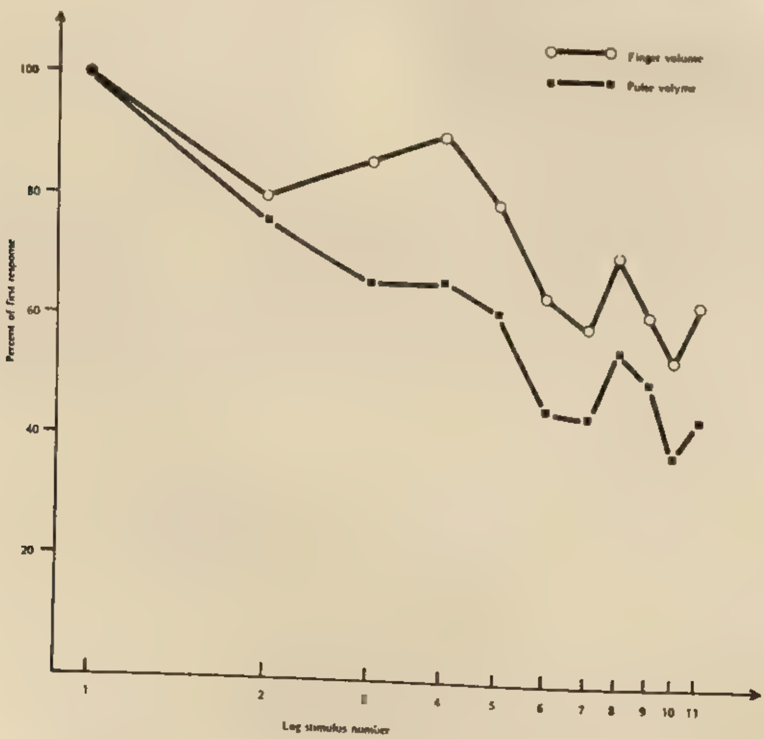


FIGURE 1. Mean responses to tones in finger volume and pulse volume, expressed as percent of the first response of each subject.

TABLE 1

t VALUES FOR MEANS OF ESTIMATES OF HABITUATION FOR FINGER VOLUME AND PULSE VOLUME (N = 19)

Measure	Finger volume			Pulse volume		
	M	SD	t	M	SD	t
H score	−1.41	2.69	−2.28**	−42.44	24.38	−7.59****
Parametric beta	−1.41	3.49	−1.76*	−42.44	35.26	5.25****
tau	1.192	.309	−2.71***	− .426	.309	−6.01****
Nonparametric beta	− .120	.388	−1.35	− 5.21	5.47	4.15****

* $p < .10$.
** $p < .05$.
*** $p < .01$.
**** $p < .001$.

There was a significant difference between the mean of Responses 2–5 and 8–11 (t = 3.43), and also significant habituation estimated by the Wilcoxon test (T = 28) for FV.

Of the five habituation estimates used only tau can be used for direct comparisons between FV and PV in regard to difference in habituation rate. A t test for correlated means showed that PV responses habituated significantly faster than FV responses, $t = 18.1$, $p < .001$.

Correlations among the five habituation estimates, first response obtained, and estimated first response (from the H-score procedure) are shown in Table 2 (FV) and Table 3 (PV). The correlations ranged from .66–.95 for FV and from .49–.92 for PV.

In Table 4 intercorrelations between FV and PV measures including estimates of habituation and obtained first response are given. The correlations ranged from .38–.73, with the correlations obtained between identical habituation estimates being generally slightly higher.

DISCUSSION

Mean habituation was highly significant for PV for all habituation estimates. For FV the result of the statistical analysis was inconclusive. As judged from the statistically most appropriate habituation estimate, nonparametric beta, FV responses did not habituate in this study. However, significant habituation for FV was obtained for all the other habituation estimates used except for the simple criterion of three successive nonresponses. In line with the findings of Davis et al. (1955) there was in the present data a significantly faster habituation for PV than for FV. However, in contrast to Davis et al., FV responses habituate as judged from most habituation estimates including that used by Davis et al. (Wilcoxon test). This discrepancy might be due to differences in the recording system. The transducer system in the present study ensures optimal reproduction of the human pulse wave (Cronvich & Burch, 1969). The

TABLE 2

PRODUCT-MOMENT CORRELATIONS AMONG ESTIMATES OF HABITUATION AND MEASURES OF FIRST RESPONSE FOR FINGER VOLUME (N = 19)

Measure	2	3	4	5	6	7
1. H score	.77	.82	.70	.66	−.03	0
2. Parametric beta		80	.75	.91	60	−.64
3. Product-moment r			.95	.75	− 24	−.27
4. Kendall's tau				.72	− 23	−.33
5 Nonparametric beta					−.49	−.62
6. First response						.91
7. Estimated first response (from H-score procedure)						—

TABLE 3

PRODUCT-MOMENT CORRELATIONS AMONG ESTIMATES OF HABITUATION AND MEASURES OF FIRST RESPONSE FOR PULSE VOLUME (N = 19)

Measure	2	3	4	5	6	7
1. H score	.69	.55	.49	.74	-.05	0
2 Parametric beta		.86	.90	.75	−.63	−.72
3. Product-moment r			.92	.70	−.59	−.67
4. Kendall's tau				.88	−.60	−.77
5. Nonparametric beta					− 47	−.58
6. First response						.83
7. Estimated first response (from H-score procedure)						—

TABLE 4

Intercorrelations among Estimates of Habituation for Finger Volume and Pulse Volume

Pulse volume	Finger volume 1	2	3	4
1. H-score	.66	.38	.41	.01
2. Kendall's tau	.49	.73	.60	−.25
3. Nonparametric beta	.65	.59	.60	−.18
4. First response	−.16	−.36	−.30	.22

disadvantages of the pneumoplethysmographic technique, e.g., sensitivity for movements of the hand and for leakage in the pneumatic system, can easily be overcome as such artifacts are readily identified in the records (Lader, 1967). Impedance plethysmography as used by Davis et al. is an indirect method.

Lader (1965) did not find significant habituation for PV using *t* test between mean size of first four and last four responses as habituation estimate. This finding might in part be explained as an effect of the type of response measure used by this author. A PV response was defined as the difference between the *smallest* pulse volume deflections occurring within 10 sec. of the onset of the stimulus and the *average* of six consecutive beats immediately before the stimulus. In that way, statisically, some of the responses obtained may be artifacts.

Increased drowsiness as a monotonous experimental situation proceeds may result in dishabituation and an uneven course of the curves. In the study by Lader (1965), which included a rest and a tone stimulation period, the pulse volume responses diminished at first, but they tended to increase again; this increase was accelerated by administration of barbiturates. Habituated and small cardiovascular responses tend to reappear and become large during drowsiness and sleep (McDonald et al., 1964; Pampiglione & Ackner, 1958). In the present study, the subjects were tested in the late afternoon, after a day's work, which might have made them prone to drowsiness. It is noteworthy that despite this the estimates of mean habituation in PV used in the present study were highly significant.

Davis et al. (1955) found that response latency was greater for FV than for PV. Operationally the two variables are defined in terms of different frequency spectrum, i.e., bandwidth, regardless of the fact that the time derivate ($\dot{v}$) of very large FV responses might approach the corresponding value for ordinary PV deflections. Electronically the two measures are separated by means of low pass (FV) and high pass (PV) filters. Such filters introduce delays of transient signals, different for different frequency bands. Thus a recorded difference in response latency between [illegible] and PV, is, at least in part, an artifact, [illegible] duced in the electronic system. For [illegible] large responses and the high frequency li[illegible] used in the present study this delay was 2 [illegible] Thus, although a highly significant difference in response latency was found between FV and PV, FV response latency being greater this finding cannot be interpreted in physiological terms. Davis et al. (1955), who found a similar difference between FV and PV, have interpreted their findings as a physiological difference.

Physiologically it is reasonable to regard FV as reflecting the tone in the venules and small veins (capacitance vessels) and PV as reflecting the tone in the arterioles and precapillaries (resistance vessels). One might expect a dependency between these two measures both on a blood flow dynamics and a neural basis. In the finger skin vessels, both for capacitance and resistance vessels, there are exclusively vasoconstrictive sympathetic fibers with noradrenaline as the transmitter (Shepherd, 1963).

The correlation between FV and PV responses was significant but moderately high. There were great interindividual differences with regard to intraindividual correlations, ranging from −.22–.98. The interpretation of these differences is uncertain. The strictly defined mechanical method of evaluation of responses used, which did not allow for adjustments due to, e.g., spontaneous fluctuations may be a source of error variance. However, a more flexible evaluation of the responses would probably have diminished the reliability of the measures.

Within each system the correlations among the five different habituation estimates are sometimes surprisingly low, indicating that the choice of habituation estimate profoundly influences the conclusion as to whether habituation has or has not occurred in the system.

Three of the habituation estimates used in the present study are of special interest: H score, because this measure has been widely accepted and used as a habituation estimate, and tau and nonparametric beta, because these two measures have certain statistical

advantages. Intercorrelations between FV and PV for these habituation estimates are somewhat lower than the corresponding intercorrelations within PV and FV. There is thus a certain independence, intraindividually, between the FV and PV response sequences with regard to habituation. This might be interpreted partly as an effect of the measurement procedure and partly as a physiological difference, the relative importance of which is uncertain.

It might be concluded that both FV and PV responses habituate, that PV responses habituate faster than FV responses, and that there is a certain independence between the two physiological variables, in regard to both the individual response sequence and habituation.

REFERENCES

Burch, G. E. Digital rheoplethysmographic study of the orienting reflex in man. *Psychosomatic Medicine*, 1961, **32**, 403–412.

Crider, A., & Lunn, R. Electrodermal lability as a personality dimension. *Journal of Experimental Research in Personality*, 1971, **5**, 145–150.

Cronvich, J. A., & Burch, G. E. Frequency characteristics of some pressure transducer systems. *American Heart Journal*, 1969, **77**, 792–797.

Davis, R. C., Buchwald, A. M., & Frankmann, R. W. Autonomic and muscular responses, and their relation to simple stimuli. *Psychological Monographs*, 1955, **69** (20, whole No. 405).

Frankenhaeuser, M., Fröberg, J., Hagdahl, R., Rissler, A., Björkvall, C., & Wolff, B. Physiological, behavioral, and subjective indices of habituation to psychological stress. *Physiology & Behavior*, 1967, **2**, 229–237.

Groves, P. M., & Thompson, R. F. Habituation: A dual process theory. *Psychological Review*, 1970, **77**, 419–450.

Hare, R. D. Psychopathy, autonomic functioning, and the orienting response. *Journal of Abnormal Psychology Monograph Supplement*, 1968, **73** (3, Pt. 2).

Koepke, J. E., & Pribram, K. H. Habituation of the vasoconstriction response as a function of stimulus duration and anxiety. *Journal of Comparative and Physiological Psychology*, 1967, **64**, 502–504.

Lader, M. H. The effect of cyclobarbitone on spontaneous autonomic activity. *Journal of Psychosomatic Research*, 1965, **9**, 201–207.

Lader, M. H. Pneumatic plethysmography. In P. H. Venables & I. Martin (Eds.), *Manual of psychophysiological methods*. Amsterdam: North-Holland Publishing Co., 1967.

Lader, M. H., & Wing, L. *Physiological measures, sedative drugs and morbid anxiety*. London: Oxford University Press, 1966.

Lidberg, L., Schalling, D., & Levander, S. E. Some characteristics of digital vasomotor activity. *Psychophysiology*, 1972, **9**, 402–411.

Lund, F. Oscillometri och pletysmografi. [Oscillometry and plethysmography.] In G. Hammarsten (Ed.), *Kliniska Laborationsmetoder II*. Stockholm: Svenska Bokförlaget, Norstedts, 1964.

Martin, I. Somatic reactivity. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology*. New York: Basic Books, 1961.

McDonald, D. G., Johnson, L. C., & Hord, D. J. Habituation of the orienting response in alert and drowsy subjects. *Psychophysiology*, 1964, **1**, 163–173.

Montagu, J. D. Habituation of the psychogalvanic reflex during serial tests. *Journal of Psychosomatic Research*, 1963, **7**, 199–214.

Pampiglione, G., & Ackner, B. The effects of repeated stimuli upon EEG and vasomotor activity during sleep in man. *Brain*, 1958, **81**, 64–74.

Sen, P. K. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 1968, **63**, 1379–1389.

Shepherd, J. T. *Physiology of the circulation in human limbs in health and disease*. Philadelphia: W. B. Saunders, 1963.

Unger, S. M. Habituation of the vasoconstrictive orienting reaction. *Journal of Experimental Psychology*, 1964, **67**, 11–17.

(Received July 31, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 706–709

MEMORIAL STRATEGY AND IMAGERY

AN INTERACTION BETWEEN INSTRUCTIONS AND RATED IMAGERY[1]

DONALD ROBBINS,[2] JAMES F. BRAY, JAMES R. IRVIN, AND PHILLIP S. WISE

Emory University

Two studies are reported in which memory instructions (interactive imagery, separation imagery, and rote repetition) and rated imagery values of the stimulus and response terms (high and low) were manipulated in a paired-associate task. The results indicated that interactive-imagery instruction led to superior performance in contrast to separation-imagery or rote-repetition instruction only with high-imagery cues. This interaction between memorial strategy and rated imagery is a confirmation of the "elusive" interaction expected by Paivio's conceptual-peg hypothesis.

Mental imagery has been shown to lead to very high levels of recall performance in paired-associate memory tasks (e.g., Paivio, 1969, 1971). In these studies, when rated imagery (Paivio, Yuille, & Madigan, 1968) was varied, Paivio (1969) found imagery effects for both the stimulus and response members of a noun pair, although the stimulus effect is apparently larger. In particular, Paivio (1969) and Humphreys and Yuille (1973), among others, have constructed noun–noun pairs representing a factorial combination of high and low imagery values for each member. They have found the following ordering of paired-associate performance for the 4 pair types: high–high (HH) > high–low (HL) > low–high (LH) > low–low (LL).

Studies have also been conducted where *S*s have been given specific instructions on how to memorize, using high-imagery noun–noun pairs. Substantially better recall performance has been found for *S*s given interactive-imagery instructions in contrast to those given rote-repetition (Bower & Winzenz, 1970; Schnorr & Atkinson, 1969; see also Paivio, 1971, for an excellent review) or separation-imagery instructions (Bower, 1970).

Thus, we have 2 classes of studies: those varying rated imagery of the stimulus and response members of noun–noun pairs and those manipulating the instructional set (or memorial strategy) using pairs of nouns high in rated imagery. In order to determine if the same processes are operating in both types of studies, there is a need to study the possible nteraction between the two. As a result, imagery effects may be better understood if we manipulate memorial strategy instructions as well as rated imagery values. Paivio (1971) studied the interaction between rated imagery and imaginal and verbal mediators. He reported mixed success in finding a statistically significant interaction between instructional set (imaginal, verbal mediation, and rote repetition) and imagery values (noun pairs of high or low rated imagery).

EXPERIMENT I

In Experiment I, 3 groups of *S*s were given either interactive-imagery, separation-imagery, or rote-repetition instructions in memorizing a list of paired associates. The list contained pairs in which the stimulus and response members were a factorial combination of high and low imagery values.

Method

Subjects. Sixty *S*s, obtained from the introductory psychology *S* pool, partially fulfilled a course requirement by participating in a 45–60-min. session.

Material and design. Each *S* was given a list of 20 noun–noun pairs using a paired-associate study–test procedure. Four types of pairs were used, in which the stimulus and response members of a pair were of HH, HL, LH, or LL imagery values, re-

---

[1] This research was supported by Grant MH-20331 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Donald Robbins, Department of Psychology, Emory University, Atlanta, Georgia 30322.

spectively, based on the Paivio et al. (1968) norms. Five pairs of each type were used, yielding a 20-pair list. The mean values on the imagery scale were 6.72 and 2.16 for high- and low-imagery nouns, respectively. It should be noted that for the material used in this study the correlation between rated imagery and rated concreteness was .97, and that between rated imagery and rated meaningfulness was .90.

The *S*s were randomly assigned to 1 of 3 groups and were instructed to memorize the noun pairs by either overt rote repetition of the noun pairs (RR), constructing an interactive visual image (II), or constructing an image of the objects denoted by the nouns in a noninteracting and separated "visual space" (SI), after Bower (1970). There were 20 *S*s in each group.

Procedure. Each *S* was tested individually. A paired-associate study–test procedure was used. On each study trial, *S* was shown a sequence of 20 noun pairs at a 6-sec. rate with a Kodak Carousel projector. On these study trials, *S* was to memorize the noun pairs in accordance with the instructions given at the beginning of the study. On study trials, the nouns were adjacent to each other, separated by a dash. A study trial was then followed by a test trial, on which the stimulus member of each pair was presented at a 6-sec. rate on the left-hand side of the slide, followed by a dash. The responses were made vocally by *S*. No information was given to *S* on the test trials. Study and test trials were alternated until *S* met a criterion of 3 consecutive successful recalls of the entire list. For each *S* on each trial, the list was presented in a different random order on both study and test trials with the restriction that 3–5 items intervene between a study trial of a particular pair and a test of that pair (and vice versa). After reaching criterion, *S* was engaged for the next minute; generally, *S*s were asked how they memorized the noun pairs. They were then told to recall all of the words in the study in any order.

Results

Since much of the research investigating instruction or strategy effects are single-trial studies, we will report first-test-trial data primarily. The mean proportions of correct responses on the first test trial for all 3 groups for the 4 pair types are shown in Figure 1. For the HH and HL pairs, Group II showed higher levels of recall than either Group RR or SI, with little differences between the 3 groups for LH and LL pairs. In addition, recall performance for Group II was ordered HH > HL > LH > LL, while that for Groups SI and RR was ordered HH > HL = LH > LL.

These observations were confirmed by an analysis of variance (as shown in Table 1) that revealed significant effects due to stimulus and response imagery and the Group × Stimulus Imagery interaction. This overall analysis was partitioned into 2 orthogonal components—one comparing Groups SI and RR, and the other comparing Group II with Groups SI and RR combined. The analysis comparing Groups SI and RR revealed significant effects of stimulus imagery, $F\ (1,\ 38) = 27.30$, $p < .05$, and response imagery, $F\ (1,\ 38) = 13.82$, $p < .05$, but not of group, $F\ (1,\ 38) = 1.65$, $p > .05$, or any other comparison. The analysis comparing Group II to Groups SI and RR combined revealed that the group effect was not significant, $F\ (1,\ 58) = 2.93$, $.05 < p < .10$, although the 3 effects found for the overall analysis comparing the 3 groups including the interaction were. The analysis of the mean total number of errors per subject item to criterion performance was similar to that of the first test trial. Since the statistical analyses of the mean total errors resulted in the same statistical conclusions as those for the first test trial, as shown in Table 1, these data are not presented, in order to avoid redundancy. In addition, high levels of recall of the pairs were found for all con-

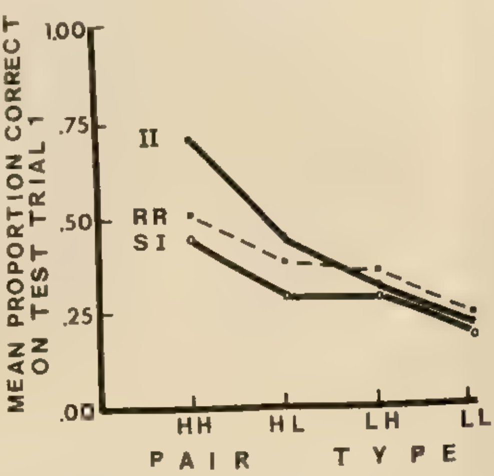


FIGURE 1. Mean proportion of correct recalls as a function of pair type (H = high imagery, L = low imagery) for the groups given interactive-imagery (II), separation-imagery (SI), or rote-repetition (RR) instructions in Experiment I.

TABLE 1

Analysis of Variance Summary

Source	*M* no. recalled on first test trial			*M* total	
	df	*MS*	*F*	*MS*	[illegible]
Between *Ss*					
Groups	2	7.63	2.34	.56	[illegible]
Ss within groups	57	3.26		1.13	
Within *Ss*					
Stimulus imagery	1	60.00	86.96**	21.96	85.1[illegible]**
Groups × Stimulus Imagery	2	4.59	6.65**	.85	3.27*
Stimulus Imagery × *Ss* Within Groups	57	.69		.26	
Response imagery	1	32.27	31.95**	10.00	75.93**
Group × Response Imagery	2	.70	< 1	.09	< 1
Response Imagery × *Ss* Within Groups	57	1.01		.13	
Stimulus Imagery × Response Imagery	1	2.82	2.78	.10	< 1
Groups × Stimulus Imagery × Response Imagry	2	.93	< 1	.06	< 1
Stimulus Imagery × Response Imagery × *Ss* Within Groups	57	1.01		.13	
Total	239				

* $p < .05$.
** $p < .01$.

ditions on the free-recall task, with little difference between groups.

Discussion

The major results of this study revealed that *Ss* given interactive-imagery instructions showed superior recall performance on the first test trial only for HH and HL pair types in contrast to *Ss* given either separation-imagery or rote-repetition instructions. These results confirm previous findings where only HH pairs have been used (Bower, 1970; Bower & Winzenz, 1970; Schnorr & Atkinson, 1969). The results of the present study also indicate that the superiority resulting from interactive-imagery instructions depends on the rated imagery value of the stimulus term. Specifically, Group II was superior in recall performance only for pairs with high-imagery stimuli. No difference due to instructions was found for pairs with low-imagery stimuli. Furthermore, the recall performance on the first test trial for *Ss* given interactive-imagery instructions was ordered HH > HL > LH > LL. The results also revealed no difference between Groups SI and RR. In addition, in contrast to Group II, the recall performance on the first test trial for Groups SI and RR was ordered HH > HL = LH > LL. Since a number of investigators have reported HL > LH when no specific instructions are given (e.g., see Paivio, 1969, 1971) we attempted to replicate the interaction found between instructions and rated imagery.

Experiment II

In an attempt to replicate the major results of Experiment I, 2 groups of *Ss* were given either interactive- or separation-imagery instructions in a paired-associate study–test procedure. Half of the noun–noun pairs were HL and the other half LH in rated imagery values.

Method

Subjects. Forty *Ss*, obtained from the introductory psychology *S* pool, partially fulfilled a course requirement by participating in a 15–20-min. session. None of these *Ss* had participated in Experiment I.

Procedure. The *Ss* were tested in small groups of up to 5 people. These small groups were randomly assigned to 1 of 2 groups ($n = 20$) and were instructed to memorize the noun pairs by either constructing an interactive visual image (II) or by using separation imagery (SI), as in Experiment I. A 46-pair list was used. On each slide, one word was at the top and the other at the bottom. Half of the noun–noun pairs were HL, while the other half were LH. On the study trial, the pairs were presented at a 6-sec. rate with a Kodak Carousel projector. The first 3 and last 3 pairs represented a primacy and a recency buffer, respectively. These pairs were not tested. The test trials were presented at a 6-sec. rate. For each pair type, half of the cue words in the test trial were high in rated imagery, and the remaining half were low in rated imagery.

As a result, the HL and LH conditions may be defined in terms of the cue word used in a test trial. Since no effect was found due to location of the

cue word (top or bottom), no further mention of this variable will be made. Each *S* recorded his responses to the cue words on a separate sheet. In addition, each small group was given a different random order of the list in both the study and test phases, and the particular cue used during testing was balanced across *S*s.

Results and Discussion

For *S*s given interactive-imagery instructions, high-imagery cues led to .64 proportion correct recalls in contrast to .45 proportion for low-imagery cues. For *S*s given separation-imagery instructions, these proportions were .33 and .29 for high- and low-imagery cues, respectively. These results were subjected to an analysis of variance which revealed significant effects of instruction, $F\ (1, 38) = 16.30$, $p < .05$, rated imagery, $F\ (1, 38) = 18.44$, $p < .05$, and a significant interaction between these effects, $F\ (1, 38) = 8.46$, $p < .05$. Thus, interactive-imagery instructions led to a large difference in recall performance for high- in contrast to low-imagery cues, while separation-imagery instructions did not. Thus, we replicated the interaction between instructions and rated imagery found in Experiment I. However, in Experiment II, Group II was superior to Group SI for both HL and LH cues, although for Group II, performance was ordered HL > LH, while for Group SI, HL = LH.

Conclusion

The finding of an interaction between instructions (memorial strategy) and rated imagery values is precisely the interaction that would be expected by the conceptual-peg hypothesis (Paivio, 1971). In particular, it appears to be a confirmation of the interaction that Paivio (1971, p. 366) has predicted but has found to be elusive. It also appears to support the "integrated structure" notion of Begg (1972), as well as the "relational association" of Bower (1970, 1973), when taken together with a retrieval axiom, namely the conceptual-peg hypothesis.

References

Begg, I. Recall of meaningful phrases. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 431–439.

Bower, G. H. Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 529–533.

Bower, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1973.

Bower, G. H., & Winzenz, D. Comparison of associative learning strategies. *Psychonomic Science*, 1970, **20**, 119–120.

Humphreys, M. S., & Yuille, T. C. Errors as a function of noun concreteness. *Canadian Journal of Psychology*, 1973, **27**, 83–94.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart and Winston, 1971.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76** (1, Pt. 2).

Schnorr, J. A., & Atkinson, R. C. Repetition versus imagery instructions in the short-term and long-term retention of paired-associates. *Psychonomic Science*, 1969, **15**, 183–184.

(Received July 31, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 710–719

EYELID CONDITIONING PERFORMANCE WHEN THE MODE OF REINFORCEMENT IS CHANGED FROM CLASSICAL TO INSTRUMENTAL AVOIDANCE AND VICE VERSA[1]

JOSEPH B. HELLIGE AND DAVID A. GRANT[2]

University of Wisconsin—Madison

Eyelid conditioning performance during Stage II instrumental avoidance reinforcement was examined after 0, 5, 10, or 20 Stage I classical reinforcements and eyelid conditioning performance during Stage II classical reinforcement was examined after 0, 5, 10, or 20 Stage I instrumental avoidance reinforcements. Response rate on the initial Stage II trial block of instrumental avoidance conditioning was a progressively increasing function of the number of Stage I classical reinforcements. Response rate on the initial Stage II trial block of classical conditioning was greater if 5, 10, or 20 Stage I instrumental avoidance reinforcements had been given than if no Stage I instrumental avoidance reinforcement had been given. Response efficiency on the initial trial block of Stage II also indicated positive response transfer from instrumental avoidance reinforcement to classical reinforcement. The results were discussed in terms of the complexities of the reinforcement and extinction processes involved in classical and instrumental avoidance conditioning of eyelid responses.

Some of the interrelationships among the responses learned under classical and instrumental avoidance reinforcement were investigated by shifting the mode of reinforcement from classical to instrumental avoidance and vice versa. As Kimble (1961, chap. 4) has pointed out, the experimental investigation of the interrelationships among the responses learned under different modes of reinforcement has proved to be extremely difficult. Distinguishing between paradigms in strictly operational terms is only an initial step in a scientific investigation of the various modes of reinforcement. Increasing attention must be given to the psychological processes involved. When note is taken of the positive and negative response transfer effects that may be observed in two-stage experiments in which the modes of reinforcement are shifted, more light may be shed on the similarities and differences of the psychological processes involved in each mode of reinforcement.

If a response learned under one mode of reinforcement can be transferred to another mode, then initial response frequency and efficiency after a shift in the mode of reinforcement should be a positive function of the number of preshift reinforcements. Grant, Kroll, Kantowitz, Zajano, and Solberg (1969) reviewed several studies which have demonstrated positive transfer from classical to instrumental conditioning and positive transfer of differential classical to differential instrumental conditioning but found no reports of positive transfer from instrumental to classical conditioning.

Grant et al. (1969) investigated transfer of responses from Stage I classical reinforcement to Stage II instrumental reward reinforcement and vice versa. They found that response rate during subsequent instrumental reward conditioning was a progressively increasing function of the number of Stage I classical reinforcements. In addition, the V or C classification of an S's eyelid responses using the Hartman–Ross (1961) criterion was usually the same in Stage I classical conditioning and Stage II instrumental reward conditioning. However, following Stage I instrumental reward training, the response rate during Stage II classical conditioning was not an increasing

[1] This research was supported in part by funds provided to the second author by the U.S. Public Health Service in Research Grant MH-06792.

[2] Requests for reprints should be sent to David A. Grant, Psychology Building, Charter at Johnson, University of Wisconsin—Madison, Madison, Wisconsin 53706.

monotonic function of the number of Stage I instrumental reward reinforcements. In fact, response rate during Stage II classical conditioning was slightly higher if no previous instrumental reward reinforcements had been given than if 20 had been given. Further, there was no significant tendency for the V or C classification of conditioned responses (CRs) in Stage II classical conditioning to correspond to the V or C classification of the responses in Stage I instrumental reward conditioning.

The classical conditioning reinforcement used by Grant et al. (1969) was a corneal airpuff—a mildly aversive unconditioned stimulus (UCS)—while the instrumental reinforcement was a symbolic reward. Before Stage I classical conditioning, their *S*s were given neutral instructions asking them to adopt a passive attitude, but before instrumental reward conditioning the *S*s were asked to adopt what amounted to an active, problem-solving attitude. The *S*s shifted from classical to instrumental reward conditioning were initially given neutral instructions and were later given the new set of instrumental reward instructions before the shift. However, the *S*s shifted from instrumental reward to classical conditioning received only the Stage I instrumental reward instructions and were not reinstructed before the shift. The asymmetry of response transfer found by Grant et al. may have been caused, at least in part, by (*a*) the asymmetry of response transfer from paradigms using aversive reinforcement to paradigms using reward reinforcement and (*b*) the asymmetry of preshift reinstruction.

Grant et al. (1969) suggested that another possible reason for the asymmetry of response transfer found by them was that the CR produced by classical conditioning was adequate to trigger the instrumental reward devices, but the response typically learned during their instrumental reward phase was of low amplitude and of a topography that would be ineffective in protecting the cornea in the subsequent classical conditioning. In one case, then, the initial classical conditioning made available a response which would be effective in the subsequent instrumental reward phase and positive, systematic transfer was obtained. In the opposite case, the initial instrumental reward training made available a response which would not be effective in the subsequent classical conditioning phase and nonsystematic, and perhaps negative, transfer performance was obtained.

Using the same transfer paradigm and avoiding some procedural problems encountered by Grant et al. (1969), the present experiment was designed to compare the transfer of responses conditioned by classical and instrumental avoidance reinforcement. The corneal airpuff can be used in both classical conditioning and instrumental avoidance training (e.g., Hansche & Grant, 1965; Kimble, Mann, & Dufort, 1955; Logan, 1951; Moore & Gormezano, 1961), so that the need for preshift reinstruction and problems with rewarding versus aversive stimuli can be eliminated. In addition, Massaro and Moore (1967) have examined the latency and recruitment of classically conditioned CRs and avoidance responses. Their analyses showed that the instrumental avoidance procedure produced a response topography similar to that of the classically conditioned CR. The implication is that the response learned during Stage I instrumental avoidance reinforcement should be effective in protecting the cornea during Stage II classical conditioning.

Method

Apparatus. The apparatus and laboratory were similar to those described by Grant et al. (1969). The *S* was placed in a soundproofed chamber. He sat in an ophthalmological examination chair facing a rectangular wooden enclosure painted flat white and illuminated to approximately 1 mL. At the back of the wooden enclosure were two ground-glass disks, approximately 10 cm. in diameter and separated from center to center by 66 cm. The disks were at eye level and approximately 120 cm. in front of the *S*. Between the ground-glass disks was a 45 × 30 cm. ground-glass screen which was not used in the experiment. The intertrial illumination of the two disks was approximately 1.5 mL. The *S* wore a plastic headset that was designed originally to support a welder's face mask. To it was attached a potentiometer to pick up the eyelid movement and a tubular jet from which the corneal airpuff could be delivered. The potentiometer was coupled

to the right eyelid of the *S* by means of a piano wire that was taped to the eyelid with plastic electrician's tape and inserted into a length of hypodermic tubing attached to the shaft of the potentiometer by a hinged coupling.

Procedure. The conditioned stimulus (CS) was the simultaneous illumination of the two ground-glass disks to a luminance of 1.9 mL. for a duration of 700 msec. The UCS, when given, was a 2 psi corneal airpuff lasting 200 msec. and delivered 500 msec. after CS onset. The intertrial interval varied from 15 to 35 sec., with a mean of 25 sec. During the instrumental avoidance training an eyelid response that produced a 1 mm. or greater deflection of the oscillograph first time-derivative pen between 200 and 500 msec. after CS onset was defined as an avoidance response (AR). An AR caused the airpuff to be omitted. A 1 mm. or greater deflection of the derivative pen corresponded to an eyelid closure of approximately 1 mm. Detection of an AR was accomplished by means of a Schmitt trigger and associated timing and logic circuitry. The puff was delivered on every trial during the classical conditioning sections of the experiment. The definition of a classically conditioned response (CR) was similar to the definition of an AR except that the scoring interval extended to 540 msec. after CS onset.

The *S*s were divided into eight experimental groups and were given neutral instructions in which they were asked not to aid or inhibit their natural eyelid responses. Four of the experimental groups (CL–AV groups) received, respectively, 0, 5, 10, or 20 classically reinforced pairings of the CS and UCS during Stage I of the experiment. They then received 60 instrumental avoidance conditioning trials during Stage II followed by 30 extinction trials with the CS alone. These four groups will be referred to as CL0-AV60, CL5-AV60, CL10-AV60, and CL20-AV60, respectively. The *S*s in the remaining four groups (AV-CL groups) received 0, 5, 10, or 20 instrumental avoidance conditioning reinforcements during Stage I of the experiment and then received 60 classical conditioning trials during Stage II followed by 30 extinction trials. These four groups will be referred to as AV0-CL60, AV5-CL60, AV10-CL60, and AV20-CL60, respectively.

A maximum of 60 trials was permitted during Stage I for *S*s in the AV-CL groups. To match groups, data from *S*s in the AV5-, AV10-, and AV20-CL60 groups were retained only if the *S* had made at least five ARs in 25 trials, because pilot experiments indicated that only *S*s who do so will typically go on to make 20 or more ARs in 60 trials. This criterion resulted in 38 rejections, approximately equally distributed among these three groups. Twelve additional *S*s were run in Group AV0-CL60 so that the data from the 12 *S*s with the lowest response rates could be rejected, resulting in an approximately equal number of rejections in each of the four AV-CL groups.

At the conclusion of the experimental session *S*s were given a questionnaire asking for their impressions of the experiment, how noxious they felt the airpuff to be, whether or not their responses were "voluntary," and whether they thought at any time that their behavior would cause the airpuff to be omitted.

Subjects. The *S*s were 80 men and 80 women student volunteers from introductory psychology courses at the University of Wisconsin—Madison. Assignment of 10 men and 10 women to each of the eight experimental groups was random until it became necessary to balance for sex.

Results

Summary of Stage I data. Table 1 gives (A) the average number of CRs or ARs, (B) the average number of trials, and (C) the average number of airpuff reinforcements during Stage I for each of the four CL-AV and the four AV-CL groups. During Stage I, the AV-CL groups received, on the average, more trials and gave more responses than the corresponding CL-AV groups. The differences between the four CL-AV groups in the mean number of Stage I airpuffs are larger than the corresponding differences between the four AV-CL groups.

Response form and transfer. All *S*s who gave one or more responses during Stage I were classified as V-form or C-form responders during Stage I, using the Hartman–Ross (1961) response-slope criterion, and were independently reclassified during Stage II. Of the 46 *S*s in the CL-AV groups who could be so classified 35 *S*s had the same V or C classification during both stages and 11 *S*s did not. Of the 60 *S*s in the AV-CL groups for whom the Stage I classification was available 50 *S*s had the same V or C classification during both stages and 10 *S*s did not. Two significant chi squares ($p < .001$), one for the CL-AV groups and one for the AV-CL groups, showed that in both cases there was a significant tendency for *S*s to have the same V or C classification in both Stage I and Stage II.

Stage II avoidance response frequency. Figure 1 shows the percent ARs given during each of the six 10-trial blocks of instrumental avoidance conditioning trials with the number of previous classical reinforcements as the parameter. The data of Vs and Cs are shown in the upper and lower panels, respectively. The final training point given for each group in Figure 1 is

the percent CRs for the final five trials of Stage I classical training. For Group CL0-AV60 the final training point is the percent ARs given on the first trial of instrumental avoidance conditioning. Since approximately 40% of the *S*s run in the AV-CL conditions were rejected on the basis of low response frequency, the 12 highest responders in the CL-AV groups are most likely to be similar to the AV-CL *S*s in terms of base response rate and conditionability. Although the results based on all CL-AV *S*s are presented, it is important to note that the pattern is identical when the data of only the high-frequency responders from the CL-AV groups are considered.

During the first 10-trial block of Stage II instrumental avoidance transfer response frequency was a progressively increasing function of the number of previous classical reinforcements, F (3, 76) = 4.95, $p < .005$. In general, during the first trial block, having had 5 or 10 previous classical reinforcements produced more responding than having had no previous classical reinforcement, and having had 20 previous classical reinforcements produced more responding than having had 5 or 10 previous classical reinforcements. Possible exceptions to this pattern are the Cs in Group CL5-AV60, but there were only four Cs in that group, so that the reliability of those data is questionable. The effect of the number of previous classical trials diminished quickly, and during the last three 10-trial blocks there was no significant effect of the number of previous classical reinforcements, F (3, 76) < 1.

Averaged over all AV trial blocks, the response rate of Vs was significantly higher than the response rate of Cs, F (1, 72) = 9.74, $p < .005$. In addition, although the initial effect of the number of previous classical reinforcements was greater for Vs than for Cs, the effect persisted longer for Cs, as indicated in a significant Number of Stage I Reinforcements × Response Form × Trial Blocks interaction, F (15, 360) = 1.75, $p < .05$.

Stage II classical conditioned response frequency. Figure 2 shows the percent CRs given during each of the six 10-trial blocks of classical conditioning with the number of previous instrumental avoidance reinforcements as the parameter. The data of Vs and Cs are shown in the upper and lower panels, respectively. The final training point is the percent ARs for the final five trials of Stage I, except for Group AV0-CL60 where the final training point is the percent CRs given on the first trial of Stage II.

The frequency of CRs during the first 10-trial block of Stage II was greater if 5, 10, or 20 ARs had been given during Stage I than if no ARs had been given, but there were no systematic differences in CR frequency between *S*s in the AV5-, AV10-, and AV20-CL60 groups during any of the 10-trial blocks. These observations were supported by a significant difference in the percent CRs on the first 10-trial block during Stage II as a function of the number of previous instrumental avoidance reinforcements, F (3, 76) = 29.26, $p < .001$, and the fact that the corresponding differences among the AV5-, AV10-, and AV20-CL60 groups were not statistically significant, F (2, 57) < 1. The response rate of the *S*s in Group AV0–CL60 reached an asymptotic level quickly, so that the number of previous instrumental avoidance reinforcements produced no significant differences in classical CR frequency during the final three 10-trial blocks of Stage II.

TABLE 1

STAGE I SUMMARY DATA

Group	A. Mean number of CRs or ARs	B. Mean number of trials	C. Mean number of airpuff reinforcements
CL0 -AV60	0	0	0
CL5 -AV60	.90	5	5
CL10-AV60	2.70	10	10
CL20-AV60	9.50	20	20
AV0 -CL60	0	0	0
AV5 -CL60	5	14.30	9.30
AV10-CL60	10	23.35	13.35
AV20-CL60	20	36.80	16.80

Note. Abbreviations: CR = conditioned response; AR = avoidance response. See text for explanation of group names.

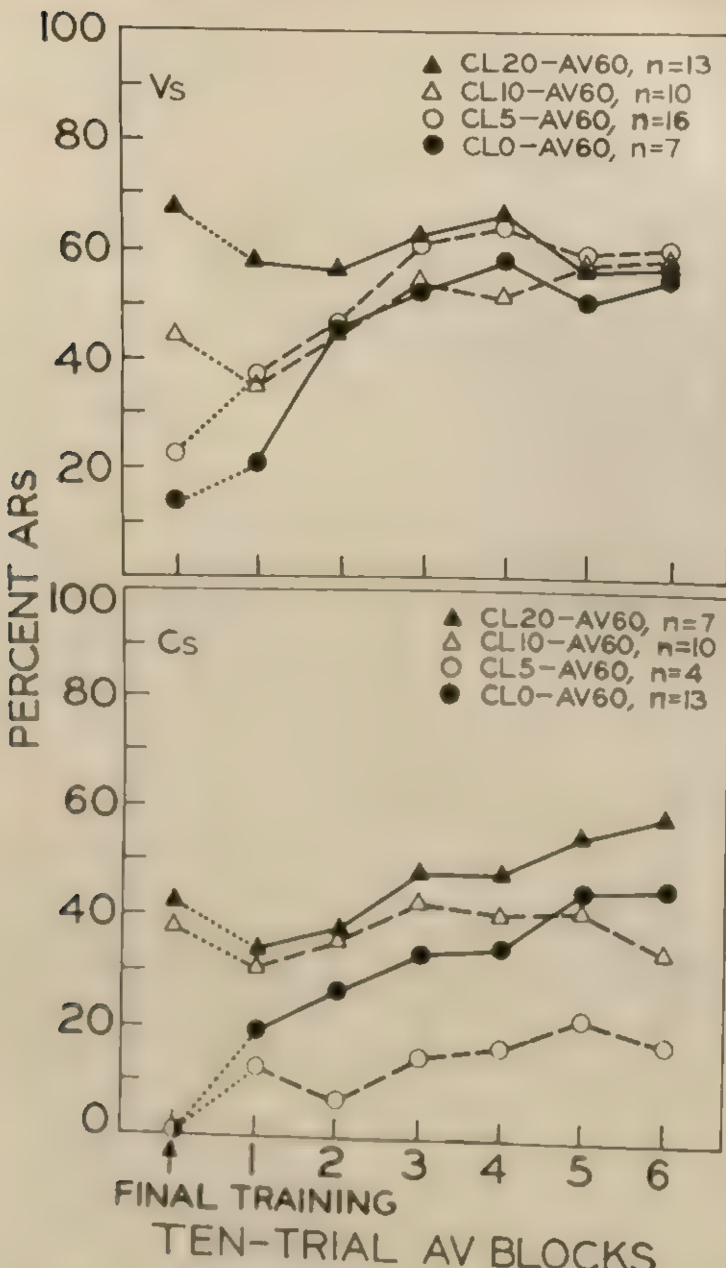


FIGURE 1. Percent avoidance re[illegible] six 10-trial blocks of Stage II [illegible] number of Stage I classi[illegible] with the [illegible] The data of [illegible] lower panel. The five trials of Stage [illegible] Group CL0-AV60 where the final training [illegible] Stage II.

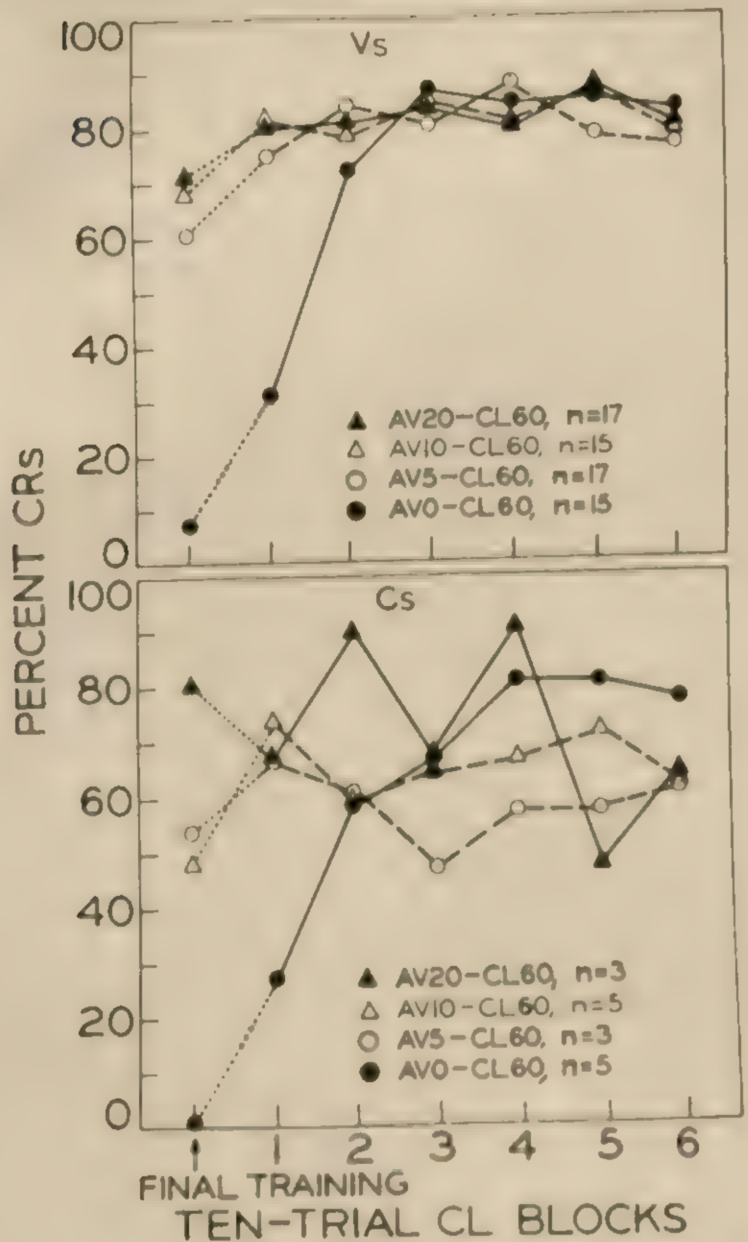


FIGURE 2. Percent conditioned [illegible] (CRs) given during each of the six 10-trial blocks of Stage II clas[illegible]itioning with the number of Stage I instrumental avoidance reinfor[illegible] the parameter. (The data of Vs are in the upper panel and the da[illegible] of Cs are in the lower panel. The final [illegible] percent CRs given on the first trial of Stage II.)

As Figure 2 shows, averaged over all Stage II trial blocks, Vs gave more CRs than Cs, $F\ (1, 72) = 12.38$, $p < .001$. In addition, the trend over Stage II trial blocks as a function of the number of Stage I avoidance reinforcements was different for Vs and Cs, as indicated by a significant Trial Blocks × Number of Stage I Reinforcements × Response Form interaction, $F\ (15, 360) = 1.75$, $p < .05$. Figure 2 also shows that only for the *S*s in Group AV0-CL60 did CR frequency increase over the successive trial blocks of classical conditioning. This is reflected in a significant Trial Blocks × Number of Stage I Reinforcements interaction, $F\ (15, 360) = 8.07$, $p < .001$.

Stage II response efficiency. Response Efficiency (PUFAMP/MAXAMP) was defined as the amplitude of eyelid closure at 500 msec. after CS onset (PUFAMP) divided by the maximum amplitude of eyelid closure between 200 and 700 msec. after CS onset (MAXAMP). During classical conditioning, MAXAMP typically occurred between 500 and 700 msec. after CS onset, making PUFAMP/MAXAMP similar to some of the efficiency measures used by Martin and Levey (1965; 1969, pp. 63–67). The mean PUFAMP/MAXAMP score was computed for each *S* on each 15-trial block of Stage II.

The mean PUFAMP/MAXAMP scores increased monotonically over the four 15-trial blocks of Stage II classical conditioning in the AV-CL groups, but did not increase over the 15-trial blocks of Stage II instrumental avoidance conditioning in the CL-AV groups. This difference between the linear trends of the AV-CL groups and the CL-AV groups was statistically significant, $F\ (1, 148) = 11.[illegible]2$, $p < .001$. Analysis of variance further indicated that the within-groups error variance of the PUFAMP/MAXAMP scores was significantly smaller during classical conditioning (AV-CL groups, error variance = .012) than during instrumental avoidance conditioning (CL-AV groups, error variance = .021), $p < .01$.

TABLE 2

MEAN PUFAMP/MAXAMP SCORES ON THE FIRST 15-TRIAL BLOCK OF STAGE II

Groups	Number of Stage I reinforcements			
	0	5	10	20
CL-AV				
Vs(n = 46)	.704	.813	.685	.796
Cs(n = 34)	.683	.838	.619	.749
AV-CL				
Vs(n = 64)	.571	.671	.747	.809
Cs(n = 16)	.444	.470	.475	.658

Note. Abbreviations: PUFAMP = amplitude of eyelid closure at 500 msec. after onset of conditioned stimulus (CS). MAXAMP = maximum amplitude of eyelid closure between 200 and 700 msec. after CS onset.

Table 2 presents the mean PUFAMP/MAXAMP scores on the first 15-trial block of Stage II for Vs and Cs in the CL-AV groups and the AV-CL groups. For the CL-AV groups the mean PUFAMP/MAXAMP scores on the first trial block of Stage II were not a monotonically increasing function of the number of Stage I reinforcements, but for the AV-CL groups the progressive increase in PUFAMP/MAXAMP scores was statistically significant, $F\ (3, 76) = 3.91$, $p < .025$.

An analysis of variance of the mean PUFAMP/MAXAMP scores for the AV-CL groups, with number of Stage I reinforcements (R), V or C response form (F) and trial blocks (B) as variables, yielded several interesting results. The Vs gave significantly higher PUFAMP/MAXAMP scores than the Cs, $F\ (1, 72) = 27.96$, $p < .001$, and their CRs showed an initially more rapid rate of PUFAMP/MAXAMP increase than the Cs, F × B interaction, $F\ (3, 216) = 3.59$, $p < .025$. The overall increase in PUFAMP/MAXAMP over trial blocks was also significant, $F\ (3, 216) = 7.57$, $p < .001$.

Extinction. In neither the CL-AV groups nor the AV-CL groups was response frequency during extinction a function of the number of Stage I reinforcements.

Questionnaire. The answers to the postexperimental questionnaire were astonishing in one respect. Of the 140 *S*s receiving some instrumental avoidance conditioning, not one reported that any of his responses would cause omission of the airpuff. The postexperimental inquiry thus confirms the reports of Massaro and Moore (1967).

Discussion

[illegible] mary results of the present study and [illegible] onship to the results of Grant et al. [illegible] 69) will be discussed in terms of their implica[illegible] regarding the nature of the processes [illegible] volved in classical and instrumental [illegible] conditioning.

[illegible] *ge II performance.* In the CL-AV [illegible] frequency during the first 10-trial [illegible] strumental avoidance conditioning [illegible] age II was a progressively increasing [illegible] ion of the number of previous classical [illegible] forcements. In addition, the *Ss* tended to retain the same V-C classification in Stage I and Stage II. These results essentially parallel the results on transfer from classical to instrumental reward reinforcement reported by Grant et al. (1969) and are consistent with the position that the exact CR learned under classical reinforcement is retained when the mode of reinforcement is changed.

Avoidance response efficiency, as defined by [illegible] IP/MAXAMP measure, was not a [illegible] ly increasing function of the number of Stage I classical reinforcements. The absence of an increasing PUFAMP/MAXAMP function at the beginning of Stage II is not [illegible] because, as Table 1 shows, the [illegible] ber of CRs given during Stage I [illegible] L0-, CL5-, and CL10-AV60 was 0, .9, [illegible] .7, respectively. When groups do not differ appreciably in the small number of CRs that have been given, systematic differences in topography measures should not be expected. Perhaps maximization of the PUFAMP/MAXAMP ratio depends upon the occurrence of enough CRs so that processes such as exteroceptive and proprioceptive feedback can influence response timing mechanisms (Grant, 1972), although Martin and Levey (1969, pp. 93–98) argue that the occurrence of overt responses may not be necessary for the development of response topography.

In the AV-CL groups, CR frequency during the first 10-trial block of Stage II was greater if 5, 10, or 20 instrumental avoidance responses had been given during Stage I than if no avoidance trials had been given, but CR frequency was approximately equal in the AV5-, AV10-, and AV20-CL60 groups during all six trial blocks of Stage II classical conditioning. Table 1 reveals a possible reason for these surprisingly equal response levels during Stage II. The CL-AV groups received 0, 5, 10, or 20 airpuff classical reinforcements during Stage I classical training. During Stage I, the AV-CL groups also received airpuffs which have long been regarded as intermittent classical reinforcement (Logan, 1951). In fact, several investigators have suggested that the occurrence of the airpuff is the primary reinforcing event in instrumental avoidance eyelid conditioning (e.g., Gormezano, 1965; Kimble, 1961, 1964; Logan, 1951). The AV0-, AV5-, AV10-, and AV20-CL60 groups received on the average 0, 9.3, 13.35, and 16.8 airpuff reinforcements, respectively, during their instrumental avoidance training. The differences in mean number of Stage I airpuffs among the AV5-, AV10-, and AV20-CL60 groups were, therefore, far smaller than the corresponding differences in the CL-AV groups. The smaller differences between the AV-CL groups in the number of such Stage I classical reinforcements could readily account for the smaller differences in the Stage II response frequencies of these groups.

The finding that *Ss* in the AV-CL groups tended to retain the same V-C classification when the mode of reinforcement was changed indicates that the same response learned during Stage I instrumental avoidance conditioning continued to be performed during Stage II classical conditioning. Positive response transfer was also indicated by the finding that the efficiency of CRs, as defined by the PUFAMP/MAXAMP measure, on the first trial block of Stage II classical conditioning was a monotonically increasing function of the number of Stage I ARs. This result also suggests that the response topography learned during instrumental avoidance training was very effective in attenuating the noxiousness of the airpuff on subsequent classical conditioning trials which may explain why the same response was retained.

In many important respects, the results in the AV-CL groups of the present experiment do not parallel the results on transfer from instrumental reward to classical reinforcement reported by Grant et al. (1969). In the Grant et al. experiment there was no indication that the response learned during instrumental reward reinforcement was the same as the CR that occurred during subsequent classical conditioning. Some or all of the differences noted in the introduction between the two experiments may account for this difference in results. The most important difference may be that the response topography learned during instrumental avoidance conditioning is effective in attenuating the noxiousness of the corneal airpuff in classical conditioning, whereas the

response topography learned during instrumental reward conditioning is less effective in this respect (Grant et al., 1969). The combined results of the present experiment and the experiments of Grant et al. suggest that when the response learned under one mode of reinforcement is "effective" under a second mode of reinforcement the response will continue to be given when the mode of reinforcement is changed from the first to the second.

Instrumental reward and avoidance reinforcement compared. In the present experiment, the AR frequency curves for the CL-AV groups converged to a common asymptote during the last three 10-trial blocks of instrumental avoidance conditioning. Grant et al. (1969) found that the instrumental response frequency curves for their four classical-to-instrumental reward groups remained parallel throughout the 60 trials of Stage II instrumental reward reinforcement and, therefore, reached different asymptotes. Differences between the relationship of Stage II reinforcement to the occurrence of a conditioned eyeblink in the two studies may account for the different patterns of results on the later Stage II trial blocks.

Grant (1973) has noted that opposing reinforcement processes occur during instrumental avoidance conditioning. The occurrence of the airpuff on no-response trials constitutes a classical reinforcement, which increases the probability of a CR on subsequent trials. The omission of the airpuff on the occurrence of an AR constitutes both an instrumental avoidance reinforcement, which increases response probability, and a classical extinction trial, which decreases response probability. Therefore, as the response rate increases, the number of classical extinction trials increases which should be expected to reduce the response rate. As a conditioning session progresses, these conditioning–extinction processes tend to produce a stable equilibrium in response rate. Therefore, even though the four CL-AV groups of the present experiment differed in initial Stage II response rate, the interaction of the processes just noted acted to produce convergence toward a common asymptote.

In instrumental reward conditioning, the number of reinforcements is a positive function of response rate. The higher the initial Stage II response rate, the higher the number of instrumental reward reinforcements. In general, higher proportions of reinforced trials should tend to produce higher asymptotic response rates (e.g., Moore & Gormezano, 1961). Therefore, the Stage II instrumental response frequency curves of Grant et al. (1969) increase in a more parallel fashion as a function of the reinforcements of the Stage II trial blocks. Thus, the differences between the results of the present experiment and that of Grant et al. can readily be understood in terms of differences between the reinforcement processes involved in instrumental reward and instrumental avoidance training of the two experiments.

Response efficiency changes over trial blocks. During Stage II the mean PUFAMP/MAXAMP ratio increased monotonically over the trial blocks of classical conditioning in the AV-CL groups. This increase is consistent with the reports of Martin and Levey (1969) and Hickok (1968), who found that CR efficiency, defined by various measures of puff avoidance, increased monotonically over trials in classical conditioning. However, during the Stage II instrumental avoidance conditioning trial blocks of the CL-AV groups, progressive increases in AR efficiency were not obtained. Maximizing the PUFAMP/MAXAMP ratio is biologically adaptive during classical conditioning. Nothing can be gained, however, during instrumental avoidance conditioning by synchronizing the point of maximum AR amplitude with the time that the airpuff would occur on a no-response trial. Therefore, one might expect that more individual variation in the CR topography would be found under instrumental avoidance training than under classical reinforcement. That the topography was actually more variable under instrumental avoidance reinforcement than under classical reinforcement was revealed by the fact that the error variance of the PUFAMP/MAXAMP measure was significantly greater with instrumental avoidance reinforcement than with classical reinforcement. The implications of the results on response efficiency can be summarized by stating that they are consistent with the suggestion of Martin and Levey (1965, 1969) that only those components of response topography that are biologically adaptive are likely to be maximized during the conditioning process.

Conclusion. Experimenters have traditionally been forced to distinguish among various forms of conditioning in terms of the experimenter's operations, rather than in terms of the *S*'s psychological processes, a very unsatisfactory situation (Grant, 1964). The findings of the present experiment and those of Grant et al. (1969), however, provide some evidence for *functional* differences as opposed to operationally defined differences among class-

ical conditioning and two forms of instrumental conditioning. Differences in response topography obtained under the different training procedures make it increasingly clear that it is inaccurate and misleading to speak of *the* eyelid CR becoming conditioned under these three modes of reinforcement. In addition, differences in the consequences of shifting from classical to instrumental avoidance or classical to instrumental reward reinforcement, etc., reveal some of the variety and complexity of the reinforcement processes that characterize these three modes of reinforcement. It is likely that further experiments using the transfer paradigm and further analyses of changes in response topography, as urged by Martin and Levey (1969), will provide a more incisive analysis of conditioning and will yield more precise specification of functional as opposed to operational differences among the various modes of reinforcement by which the eyelid response can be conditioned.

REFERENCES

Gormezano, I. Yoked comparisons of classical and instrumental conditioning of the eyelid response; and an addendum on "voluntary responders." In W. F. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1965. Pp. 48–70.

Grant, D. A. Classical and operant conditioning. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964. Pp. 1–31.

Grant, D. A. A preliminary model for processing information conveyed by verbal conditioned stimuli in classical differential conditioning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning, II: Current theory and research.* New York: Appleton-Century-Crofts, 1972. Pp. 28–63.

Grant, D. A. Reification and reality in conditioning paradigms: Implications of results when modes of reinforcement are changed. In F. J. McGuigan & D. B. Lumsden (Eds.), *Contemporary approaches to conditioning and learning.* Washington, D.C.: Winston, 1973. 49–67.

Grant, D. A., Kroll, N. E. A., Kantowitz, B., Zajano, M. J., & Solberg, K. B. Transfer of eyelid conditioning from instrumental to classical reinforcement and vice versa. *Journal of Experimental Psychology*, 1969, **82**, 503–510.

Hansche, W. J., & Grant, D. A. A comparison of instrumental reward and avoidance training with classical reinforcement technique in conditioning the eyelid response. *Psychonomic Science*, 1965, **2**, 305–306.

Hartman, T. F., & Ross, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *Journal of Experimental Psychology*, 1961, **61**, 334–338.

Hickok, C. W. A study of the effects of informative and facilitatory instructions on human differential eyelid conditioning. Unpublished doctoral dissertation, University of Wisconsin, 1968.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

Kimble, G. A. Categories of learning and the problem of definition: Comments on Professor Grant's paper. In A. W. Melton (Ed.), *Categories of human learning.* New York: Academic Press, 1964. Pp. 32–45.

Kimble, G. A. Mann, L. I., & Dufort, R. H. Classical and instrumental eyelid conditioning. *Journal of Experimental Psychology*, 1955, **49**, 407–417.

Logan, F. A. A comparison of avoidance and nonavoidance eyelid conditioning. *Journal of Experimental Psychology*, 1951, **42**, 390–393.

Martin, I., & Levey, A. B. Efficiency of the conditioned eyelid response. *Science*, 1965, **150**, 781–783.

Martin, I., & Levey, A. B. *The genesis of the classical conditioned response* (International Series of Monographs in Experimental Psychology, No. 8). London: Pergamon Press, 1969.

Massaro, D. W., & Moore, J. W. Differential classical and avoidance eyelid conditioning. *Journal of Experimental Psychology*, 1967, **75**, 151–157.

Moore, J. W., & Gormezano, I. Yoked comparisons of instrumental and classical eyelid conditioning. *Journal of Experimental Psychology*, 1961, **62**, 552–559.

(Received August 2, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 720–721

SHORT REPORTS

NONLEARNING: THE COMPLETENESS OF THE BLINDNESS[1]

PAUL FINGERMAN AND MARVIN LEVINE[2]

State University of New York at Stony Brook

College students fail to learn a simple discrimination (e.g., choosing A from the pair A–B), if they are set for a different type of solution. It is possible, however, that these *S*s learn enough to recognize the solution when stated. In the present study, therefore, *S*s who failed to learn this discrimination were given a multiple-choice test containing the statement of the solution as one of the alternatives. None of the *S*s selected this alternative.

Exploring one facet of his theory of learning, Levine (1971) derived conditions for the occurrence of *nonlearning*, the absence of learning, in the simplest of tasks. The conditions of failure were specified for even a well-motivated, intelligent adult who perceived all the relevant stimuli. Before reviewing the derivation, it is worth stressing that the task, simultaneous discrimination, is almost irreducible in its simplicity. Two letters, A and B, are both printed on each of a long series of cards, varying only in which of the letters is on the right (R) side of the card and which on the left (L). The *S* turns the cards 1 at a time and is instructed to state 1 of the 2 letters at each card. The *E* says *correct* whenever *S* says *A* and *incorrect* whenever *S* says *B*. The *S*s are instructed to try to be correct as often as possible. College students normally learn this task quickly, making their final error at about the second card.

The theory, which specifies the conditions under which the choice of A will not increase in frequency, may be summarized as follows: The *S* samples from a set of hypotheses in trying to solve this problem. If he samples from an incorrect set (one which does not contain the solution) he will not solve the problem. In such a circumstance, he will "learn," i.e., solve the problem, only by first exhausting the incorrect set. When it becomes empty he samples a new set, one which (hopefully) contains the solution. An *S*, then, who starts with an incorrect set of hypotheses can, nevertheless, solve the problem. The eventual emptying of this set serves as a signal to try something new.

Suppose, however, that this initial incorrect set of hypotheses is infinitely large—it never becomes empty, and *S* never receives a signal to try another set of hypotheses. For the specific task described above, suppose *S*'s initial hypothesis set is that the solution is a complex position sequence (e.g., it *double alternates* [LLRR] *for 4 trials, then single alternates* [LRLR] *for 4 trials, then repeats*). The set of such position sequences is virtually infinite, yet does not contain the hypothesis *say A*. According to the theory, *S* should learn nothing about the contingency between *A–B* and *correct–wrong*.

One can plausibly cause *S* to sample this incorrect but vast hypothesis set by presenting him first with preliminary problems having just such position-sequence solutions. Levine (1971) performed 2 experiments in which 6 preliminary problems with various position-sequence solutions preceded the critical problem with the *say-A* solution. He found 80% of the *S*s failed to solve the final problem in 100 trials, and were saying *A* about 50% of the time throughout all these trials. The prediction seemed confirmed, then, that *S* learns nothing about the simplest contingencies if the appropriate hypothesis is not in the set.

The question arises, however, whether *S* really learns *nothing* about the relationship between *A* and *correct*. The influence of this invariant repetition might be detected by some more sensitive probe. One possibility is that *S*, though unlikely to formulate the correct hypothesis, might recognize it when stated. If, for example, the correct rule statement were included among a small set of alternatives, *S* might be able to consult his memory of a few of the trials and recognize the existence of this invariant relationship. An *S*, therefore, responding *A* at chance level (i.e., failing by the normal criterion) might nevertheless select the correct alternative from a simple multiple-choice test. The following experiment was performed to investigate this possibility.

Method. Forty-one Stony Brook undergraduates served as *S*s. The data from 9 *S*s were discarded for reasons described below. Of the remaining 32 *S*s, half were fulfilling a requirement in the introductory psychology course, and half were paid volunteers.

Two decks of 110 3 × 5 in. cards each were developed. Every card had the letters A and B printed on the center horizontal line. Within each deck, half of the cards had A on the left and B on the right; the other half had the reverse. The 2 types of cards appeared randomly within each deck, and the 2 decks differed only in the sequence of these cards.

Four additional cards were prepared, each of which contained the following statement and 6

[1] This work was supported by Research Grant MH-11857 from the National Institute of Mental Health to the second author.

[2] Requests for reprints should be sent to Marvin Levine, Department of Psychology, State University of New York, Stony Brook, New York 11790.

alternatives:

I said "correct" on this last problem:
(a) whenever you said "A".
(b) whenever you paused for more than two seconds before responding.
(c) whenever you responded according to the sequence LLRLRRLLRLLL.
(d) whenever you responded according to some other sequence.
(e) whenever you responded according to a sequence. The correct sequence, however, kept changing.
(f) None of the above.

The sequence of the alternatives was counterbalanced over the 4 cards, with the restriction that (using the above letter labels) Alternative d always followed c, and f was always last. Also, each of the 4 cards was replicated except that "B" appeared in Alternative a.

The *S*, seated across a table from *E*, was presented with a deck and instructed to go through the deck 1 card at a time, choosing on each card 1 of the 2 letters. He was further told that *E* would say *correct* or *wrong* according to some rule, and that he could be correct all of the time if he discovered the rule. This routine was then followed as *S* went through the deck. Seven problems were presented in this way. The first 6 problems had position-sequence solutions beginning with double alternation (RRLL) and ending with LLRLRRRLRR. The *S* was run in each of these 6 problems until he either made 15 successive correct responses or completed 95 trials. The *E* announced the solution at the end of each of these problems and then said, "Here is another problem," and presented the other deck. The final, critical problem also began in this style. For half of the *S*s, the solution rule was *say A*; for the other half, it was *say B*. If *S* made an error after Trial 60, *E* said: "You seem to be having some trouble with this problem. A lot of subjects seem to have trouble with it, so I've prepared a multiple-choice question I would like you to answer to help me to figure out why." He then presented the appropriate multiple-choice card and said, "What do you think is the best answer?" After *S*'s response, *E* said, "Now I'd like you to rank the answers from best to worst."

On the critical problem, 2 different decks and 2 solutions (*say A* or *say B*) defined 4 conditions. These, combined with the 2 conditions of payment and with the 4 multiple-choice sequences, produced 32 unique conditions. The *S*s were run until there was 1 *S* in each condition who (*a*) solved at least 1 of the first 6 problems, and (*b*) failed to solve the critical problem.

Results. Since the theory is best tested if *S* is motivated to be correct and if he is indeed sampling among sequence hypotheses, the protocols of any *S* who solved none of the 6 preliminary problems were eliminated from consideration. Five of the 41 *S*s were dropped for this reason. None of these 5 *S*s, incidentally, solved the critical problem.

There were, then, 36 *S*s who solved at least 1 sequence problem. Only 4 of these solved the final problem. This repeats the previous finding that the substantial majority of college students fail to learn this simple problem. For the 32 failing *S*s, the percentages of correct responses on each of the 6 successive blocks of 10 trials were 48%, 52%, 50%, 53%, 51%, and 47%.

The multiple-choice test was presented to these 32 *S*s in the search for the merest modicum of learning. The results were decisive. None of the *S*s selected the correct alternative as the solution. Twenty-seven *S*s picked 1 of the 3 sequence-solution alternatives. The remainder picked "none of the above." The *S*s had also been asked to rank order the alternatives. The mean rank of the correct alternative was 4.5. It tied for last place with the alternative concerning waiting for 2 sec. before responding, the alternative most irrelevant to the experiment. Several *S*s remarked on the unlikelihood of either of those 2 alternatives and said that they could not decide which to put last.

The conclusion seems clear. Although *S*s were forced, by stating the letter, to attend to the critical stimuli, and although the feedback was desirable, perceivable, and unambiguous, *S*s failed to register the simple invariant relationship between the two. We can assert with greater confidence that *S*s learn nothing about hypotheses not in their hypothesis set.

REFERENCE

LEVINE, M. Hypothesis theory and nonlearning despite ideal S–R reinforcement contingencies. *Psychological Review*, 1971, **78**, 130–140.

(Received December 15, 1972)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 722–725

MEMORY SCANNING:

EFFECT OF UNATTENDED INPUT[1]

MARILYN CHAPNIK SMITH[2]

University of Toronto, Toronto, Ontario, Canada

DAVID BURROWS

State University of New York College at Brockport

To determine whether both attended and unattended items are held in a common short-term memory (STM) store, a Sternberg varied set reaction time (RT) paradigm was employed. Four word lists were presented either monaurally or with simultaneous dichotic presentation of a second 4-word list which *S*s were instructed to ignore. Fifty per cent of the negative probes consisted of nonlist items, while the remaining 50% were selected from the unattended channel. The error rate for the latter probe words, which required a *no* response, was 20%, indicating that there was some STM storage for the unattended items. The RT for the positive probes in the dichotic condition increased by 40 msec. These results were interpreted as suggesting that attended and unattended items are differentially tagged or in some other way kept separate in memory, but that random sampling of the unattended channel takes place.

While early investigators found no evidence of long-term memory (LTM) for verbal items which were not attended (Cherry, 1953; Moray, 1959), several recent studies have indicated that there is some short-term memory (STM) for such unattended inputs (Glucksberg & Cowen, 1970; Norman, 1969; Peterson & Kroener, 1964). As yet little is known about the storage properties of unattended materials. One question of interest is whether attended and unattended items are kept in a common memory store.

To study this problem, Davis and Smith (1972) examined the serial position curves for lists of attended words as a function of whether words that *S*s were instructed to ignore were presented in an unattended channel. It was argued that if all items are held in a common store, the presence of the unattended materials should result in an attenuation of the recency effect observed for attended materials, making it extend over fewer serial positions. No such change in serial position was found, suggesting that information from the 2 channels was not kept in common storage.

There were, however, 2 findings which suggested that despite the absence of a change in serial position there was some memory for unattended input. First, there was an overall decrease in accuracy when the unattended material consisted of words rather than nonsense syllables. Norman (1969) argued that, to be analyzed for meaning, items must make contact with LTM store. Second, when free recall was employed some intrusion errors were found from the unattended channel. When these intrusions were plotted as a function of their input serial position, a roughly inverse exponential decay of intrusions over serial position was found, suggesting an STM decay function for the unattended items. Further, in plotting these intrusions as a function of *output* serial position, it was found that they were offered only in the later output serial positions, i.e., after the recall of several other items. This latter finding suggested that attended and unattended inputs were differentially "tagged" (Yntema & Trask, 1963), and that unattended items were offered only after their "ear tag" had decayed. These results suggest that while there is some memory for unattended items, they appear either to be differentially tagged or in some other way stored separately from the attended materials.

A different paradigm for examining the storage of unattended verbal materials is presented here. Two lists of words, each 4 items in length, were presented dichotically. The *S*s were instructed to attend to one ear and to ignore the materials presented to the other. Following presentation of the lists, a probe word was presented. A *yes* response was required if the probe was one of the words in the previously presented attended list, and a *no* response if it was not. Sternberg (1966), using single list presentation, suggested that when such a procedure is employed, *S*s do a high-speed memory scan to determine if the probe was a member of the previously presented list.

In this study, to learn something about the storage of unattended materials, *S*s were occasionally presented with probe words that had in fact been presented to the unattended ear. A negative response to such probe words was required. If these words had no representation in STM, *S*s should treat them just as they would words which had not been presented at all—i.e., error rate and reaction time (RT) should be no greater than for other negative probes. If, on the other hand, all items from both attended and unattended lists are held in common store in an undifferentiated form, *S*s should always respond *yes*—i.e., there should be an error rate of 100%. Following Sternberg's (1966) as-

[1] This research was supported by National Research Council Grant APA-191 to M. Smith. The authors wish to thank Bert Forrin for his helpful comments on an earlier draft of this paper.

[2] Requests for reprints should be sent to Marilyn C. Smith, Department of Psychology, University of Toronto, Scarborough College, West Hill, Ontario, Canada.

sumption that a serial scan of the memory set is made in order to reach a decision, twice as many items would have to be scanned in this condition. Consequently, RTs for both positive and negative probes should be considerably greater than RT in a control condition where only a single attended list is presented. A third possibility, suggested by the Davis and Smith (1972) study, is that the 2 lists may be tagged according to ear of arrival, or in some other way kept in a separate store in memory. If the tagging were perfect, and if *S*s were able to limit their scan only to the attended list, then even though the items were held in STM they should be handled in the same manner as regular negative probes. To the extent that the tagging was not perfect, however, or that there was a fading of tag information with time, some false positives would be expected, giving a greater error rate for negative probes from the unattended list. To test this possibility, word lists were presented at both fast and slow rates, with the expectation that there would be less fading of tag information, and hence fewer errors, under a fast presentation rate.

Method. The *S*s were 8 male secondary school students, all right-handed according to self-report. They were paid $2.00 per session for serving in the experiment.

The words used for each trial were sampled from a pool of 720 one-syllable words that had a rating of 45 or higher on the Thorndike–Lorge (1944) G count, and both began and ended with a spoken consonant. In cases where 2 or more words were homophones or syntactic variants of each other (e.g., give–gave), only one form of the word was retained in the pool. Only nouns, verbs, and adjectives were used.

The lists used in the experiment were generated by computer program and recorded by a male voice using a Sony TC-355 2-channel tape recorder. The *S*s heard the lists through Sharpe HA-10 headphones and responded by pressing one of 2 push buttons. The *S* held one button in each hand. The RT for each response was measured to the nearest millisecond by a Hunter Klockounter. Each *S* was run individually in a quiet room. A plywood partition separated *S* and *E*.

On each trial, *S* heard a warning tap, a list of 4 memory words, another warning tap, and a probe word. The *S* was required to decide whether or not the probe word was present in the list and to make a response as quickly as possible. In the monaural condition, *S* heard all stimuli through one ear only. For any given *S*, this was the same ear throughout the experiment, and was designated as the "relevant" ear for that *S*. In the dichotic condition, words were also presented to the irrelevant ear. Each irrelevant word was presented at approximately the same time as a memory word. Nothing was presented to the irrelevant ear during presentation of warning signals or probe word.

In the dichotic condition there were 3 types of probe words: (*a*) *Positive probes* were words that had been presented to the relevant ear as part of the memory set. The *S*s were instructed to press the *yes* button in response to positive probes. These probes were taken from each of the 4 serial positions equally often. Fifty per cent of all probes were positive. (*b*) *Negative intralist probes* were words that had been presented to the irrelevant ear while the memory set was being presented to the relevant ear. The *S*s were instructed to press the *no* button in response to these words. Twenty-five per cent of the probes were of this type. They were drawn from each serial position of the irrelevant word set equally often. (*c*) *Negative extralist probes* were words that had never been presented to either ear. They were always treated as *no* words. The *S*s were made fully aware of the nature of the *no* words and were especially encouraged to block out the irrelevant ear in the dichotic condition.

In the fast condition, the memory set, warning signals, and probe words were presented at the rate of 2 items per second. In the slow condition, the rate was 1 item per second. Each *S* served for 1 practice day and 4 test days. Each test day consisted of a monaural tape and a dichotic tape at one of the 2 rates. Each tape consisted of 4 practice trials followed by 80 test trials (40 *yes*, 40 *no*) presented in a random order. There were a total of 8 tapes, all dichotic. The monaural conditions were created by turning off the switch for the irrelevant channel. Half of the *S*s heard each tape as a monaural tape, and half heard each as a dichotic tape. Assignment of response type to button and order of serving in each Mode × Rate combination was balanced across *S*s.

Results. The error rates are shown in Figure 1. In the monaural condition, error rates for both *yes* and *no* responses are a relatively low 4% (Figure 1, Section a). In the dichotic condition, the error rate for the *yes* and external *no* responses does not differ from that found in the monaural condition. However, when the probe was a word that had been presented in the unattended ear (internal probe), the error rate was very high, jumping to 20%. Most of these errors were due to *S*'s failure to make a correct rejection—i.e., a false positive response was made (Figure 1, Section b). In Figure 1 Section c, these internal errors are plotted as a function of their serial position in the unattended ear. Most of the errors occurred at the middle 2 positions.

All RT analyses were based on correct responses only. Figure 2 shows mean RT for dichotic and monaural conditions as a function of serial position. Dichotic trials were on the average 42 msec. slower than monaural trials, $F\ (1,\ 7) = 5.79$, $p < .05$. There was also a significant effect of serial position, $F\ (3,\ 21) = 4.48$, $p < .05$. For both monaural and dichotic conditions, the serial position curve is bowed. The shortest RTs were for probes of the last position, followed by probes of the second-to-last position.

A repeated measures analysis for the RT data on negative trials revealed a significant effect of monaural vs. dichotic conditions, $F\ (1,\ 7) = 5.76$, $p < .05$, and no other significant effects.

Discussion. This study was conducted to determine whether unattended words are placed in the

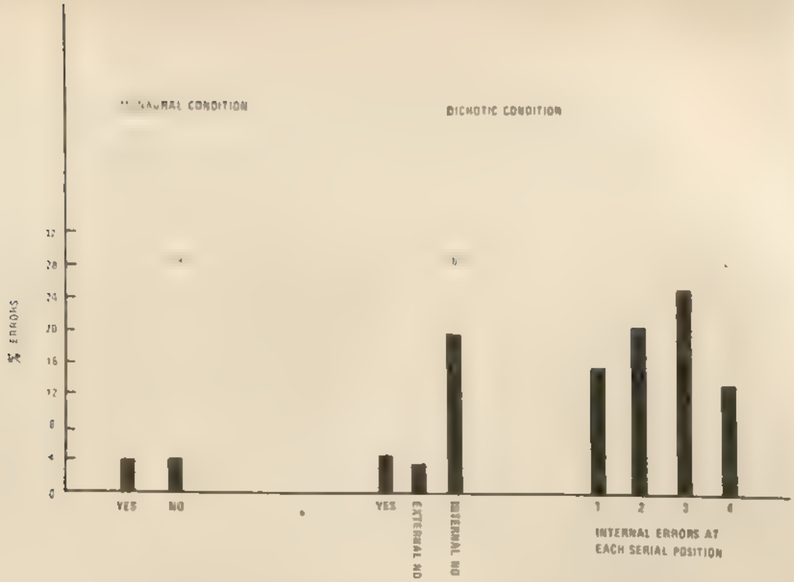


FIGURE 1. *Yes No* error rates for monaural and dichotic conditions.

same STM store as those to which *Ss* are paying attention. Clearly, there is some differentiation in STM between attended and unattended inputs. If both materials were simply placed in a common store, with *S* being required to scan all the items in order to reach a decision about the probe, all probes from the unattended channel would have been treated as positive items. This would have yielded an error rate of 100% for intralist negative probes. However, the error rate found was 20%. Second, given Sternberg's (1966) estimate of 38 msec. to scan an item in memory, adding 4 items in the dichotic condition would have increased RT by about 150 msec. Instead, an increase of approximately 40 msec. was found. Both the error and the latency data thus indicate a differentiation in memory between attended and unattended items.

Why, then, is there such a high error rate for intralist negative probes? One possibility is tha although the items in the 2 channels are differentiall tagged according to ear of input, there is a fadi of the tag information with time. Consequently, when a probe is presented which was an item in the unattended channel, *Ss* will respond *yes* if the item is still available in memory but the tag information has faded away. However, this hypothesis would have predicted fewer errors with a fast presentation rate, since there is less time for fading. Yet no effect

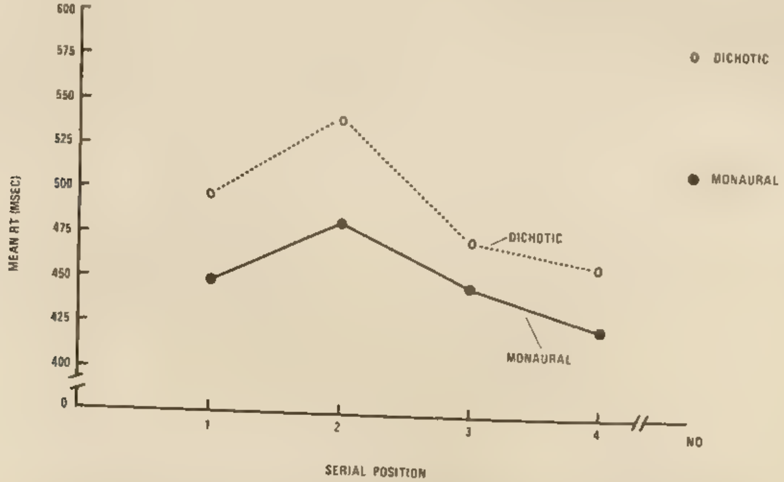


FIGURE 2. Mean reaction time (RT) for positive responses under monaural and dichotic conditions as a function of serial position. (Negative responses are shown on the right.)

of rate was found. Second, we would have expected the greatest number of errors when the probe was an item from the beginning of the list, with the number declining as the end of the list is approached. Instead, a bow-shaped serial position curve was found for errors, with the maximum number of errors occurring in the middle 2 positions.

A second possibility is that all presented items are placed in a common store, but that due to attenuation of the unattended input (Treisman, 1964), the trace strength of their memorial representations is very weak. We must, therefore, assume that the decision as to whether the probe was presented in the memory set is not based upon a serial scan but instead depends on trace strength. If the trace strength is not fixed, but varies about some mean, then on some trials it will be sufficiently high for *S*s to make a positive response to the probe. To account for the data, we must assume that trace strength for unattended materials does not vary greatly over serial position and is not sensitive to the different rate presentations employed here. This would account for the higher error rate for intralist than extralist negative probes. To explain the overall latency increase in the dichotic condition, we must postulate that the addition of these low-strength traces in memory adds noise, thereby causing a decrease in efficiency.

Finally, a different mechanism which could handle these data is one in which an attention switch to the unattended channel occurs during the list presentations, thereby allowing for one of the unattended items to be stored with the attended items. Time of occurrence for such a switch would be random, but would have increased likelihood of occurring in the middle of the list. If such a switch occurs once, permitting one additional item to enter the attended store, *S*s in the dichotic condition would have one more item to scan. The value of the increase in RT of 42 msec. found in the dichotic condition is just what would be expected for the scanning of an additional item. Further, if one item from the list of 4 does get tagged as an attended item, an error rate of 25% would be expected—a value which is very close to the 20% which was in fact found.

It is important to note that if this attention switching did take place, it must have occurred involuntarily, for *S*s were instructed to ignore the material in the irrelevant ear, and there was no reason for *S* to sample the unattended input in order to perform the task. On the contrary, his performance would be better if this switching did not occur at all.

Such an interpretation has several difficulties. It makes predictions opposite to the findings of Davis and Smith (1972), for such a mechanism must postulate a change in serial position data for the attended input. Second, it would bring into question the hypothesis of STM for unattended verbal items: it would suggest that items in the unattended channel reach STM only when an involuntary attention switch occurs, allowing them to pass into STM. This would seem unlikely, in view of the many recent studies in which evidence for STM for unattended materials has been found.

REFERENCES

CHERRY, E. C. Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 1953, **25**, 975–979.

DAVIS, J., & Smith, M. C. Memory for unattended input. *Journal of Experimental Psychology*, 1972, **96**, 380–388

GLUCKSBERG, S., & COWEN, G. N., JR. Memory for nonattended auditory material. *Cognitive Psychology*, 1970, **1**, 149–156.

MORAY, N. Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 1959, **11**, 56–60.

NORMAN, D. Memory while shadowing. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 85–93.

PETERSON, L., & KROENER, S. Dichotic stimulation and retention. *Journal of Experimental Psychology*, 1964, **68**, 125–130.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

TREISMAN, A. M. Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 449–459.

YNTEMA, D. B., & TRASK, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 65–74.

(Received January 13, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 726-728

RETRIEVAL OF REPEATED ITEMS EMBEDDED IN CHANGING LISTS[1]

WALTER RITTER[2]
Herbert H. Lehman College, City University of New York

HERMAN BUSCHKE
Albert Einstein College of Med.

MARION GINDES
Brooklyn College, City University of New York

A set of target words from the same category, embedded in lists also containing other items, was repeated across all trials for free recall of all items presented on each trial. Retrieval of the repeated items from long-term storage was less when other items, from a different category, were new on each trial rather than also repeated across trials; and retrieval was still less when the other new items were from the same category as the repeated items. Long-term storage of the items in the repeated set was equivalent in all 3 conditions, confirming that retrieval of items within a list from long-term storage is affected by the context provided by the rest of the list.

It has been pointed out (e.g., Bower, 1972; Kintsch, 1970) that the properties of a list of items as a whole affect the learning of individual items in the list. This view can be examined by presenting a set of words as part of larger lists containing other words, so that the target words would be learned in different contexts, i.e., as parts of different lists. Accordingly, half of the items in a multitrial free-recall experiment were repeated from trial to trial, while the rest of the items were new on each trial. Since *S*s were required to recall all of the items presented on any given trial, their task was to learn the repeated items across trials, while also recalling the new items for only 1 trial.

The set of repeated items were all from the same category (BIRDS), the new items being drawn from the same category (BIRDS) for one group of *S*s and from a different category (TREES) for another group of *S*s. In addition, a third group of *S*s was given only repeated items, so that all items could be learned over trials; half of the items were the repeated BIRDS given to the other 2 groups and the rest were a set of TREES also repeated on all trials. The context in which the repeated subset of items (BIRDS) was presented to all 3 groups, therefore, was varied in 2 ways: (*a*) the balance of the items were from either the same or different categories, and (*b*) with same-category items, the context was nonrepeated, whereas with different-category items, the context was either repeated or nonrepeated. Finally, all 3 experimental groups were given an additional trial in which the repeated BIRDS were presented by themselves for free recall, and a control group was also given the repeated BIRDS alone, but without any previous trials.

Method. Eight successive lists of 16 items each were read aloud at the rate of 1 item/sec for free recall of each list immediately after presentation. There were 3 conditions, and the same 8 BIRDS were presented on all trials in each. On each trial, the repeated BIRDS were alternated with the other 8 items to distribute the 2 classes of items evenly over the list. From trial to trial, the order of the repeated BIRDS was randomized over their designated serial positions. In Condition B-b, the 8 repeated BIRDS alternated with the names of 8 BIRDS which were new on each trial. In Condition B-t, the repeated BIRDS alternated with a new set of TREES on each trial. In Condition B-T, the repeated BIRDS alternated with 8 TREES, which were repeated across trials. These repeated TREES were randomized within their designated serial positions across trials.

On each free-recall trial, *S*s wrote the items that were recalled on a separate sheet of paper. The *S*s were told about the structure and content of their lists, and were given 90 sec. for recall after each list was read. The 8 repeated BIRDS were presented alone on Trial 9 to all experimental *S*s and for only 1 trial to 20 control *S*s, who recalled them for the first time.

The *S*s were undergraduate students between the ages of 19 and 25 enrolled in psychology courses at Lehman and Brooklyn Colleges. A total of 80 *S*s participated in the experiment, 20 *S*s being randomly assigned to each of the 3 experimental conditions and to the control group.

The relative contributions of short-term storage (STS) and long-term storage (LTS) to total recall were estimated by the method of Tulving and Colotla (1970). Recall from short-term storage includes those items with 7 or less other items intervening between presentation and recall, counting both presented and recalled items. Retrieval from LTS includes those items with more than 7 intervening items between presentation and recall. The number of items retained in LTS is the cumulative number of items retrieved from LTS at least

[1] This research was supported by U. S. Public Health Service Grant MH-17733 to the second author from the National Institute of Mental Health, and Grants NB-03356 from the National Institute of Medical Science and HD-01799-07 from the National Institute of Child Health and Development. The authors thank Gerald Lazar for critical discussion.

[2] Requests for reprints should be sent to Walter Ritter, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 14061.

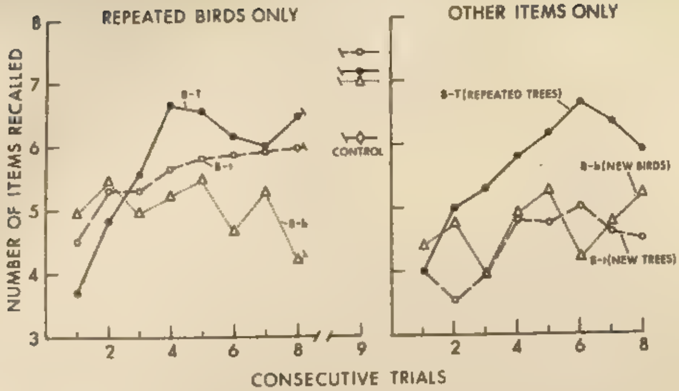


FIGURE 1. The left panel presents the number of repeated BIRDS recalled when presented for recall together with new BIRDS, new TREES, and repeated TREES, with results for Trial 9 representing recall of repeated BIRDS when presented by themselves; the right panel represents concurrent recall of new BIRDS, new TREES, and repeated TREES.

once, on the assumption that once an item has been retrieved from LTS, it remains in LTS. In support of this assumption, recent work has shown that when items on a list are no longer presented after they have been recalled just once (i.e., only items on the list never recalled are presented on subsequent trials), 85% will be spontaneously retrieved subsequently, even after several intervening trials where they have not been recalled (Buschke, in press).

Results. The left panel of Figure 1 shows the recall across trials of the repeated BIRDS embedded in the lists of the 3 conditions, and the right panel shows the recall of the rest of the items presented for recall on each trial in each of the 3 conditions. Recall of the repeated BIRDS increased most over trials (sum of Trials 1–4 vs. sum of Trials 5–8) in Condition B–T, t (19) = 4.98, $p < .001$; increased moderately in Condition B–t, t (19) = 3.52, $p < .01$; but did not increase at all in Condition B–b. There was a significant effect of conditions on the recall of the repeated BIRDS on Trial 8, F (2, 57) = 3.40, $p < .05$. The *S*s in Condition B–t had a greater overall recall of the repeated than of the new items, t (19) = 4.58, $p < .001$, whereas *S*s in Condition B–b did not differ in their overall recall of the repeated vs. the new items.

Evidence that the failure of the B–b group to increase in recall of the repeated BIRDS across trials was due to difficulty in retrieval from, rather than storage in, LTS was obtained on Trial 9, where the repeated BIRDS were presented alone to all groups. While there was no significant difference in recall among the 3 experimental conditions, recall by the control group, who were recalling the repeated BIRDS for the first time, was significantly lower, F (2, 76) = 8.54, $p < .01$. If the B–b group had not benefited from exposure to the repeated items during Trials 1–8, then they should not have done better on Trial 9 than the control group did on their single trial. Any benefit from previous trials presumably must be mediated by LTS.

To further substantiate the contention that the lower recall scores of the B–b group were due to less effective retrieval from LTS rather than less effective storage in LTS, estimates of retention in LTS, retrieval from LTS and recall from STS across trials are presented separately in Figure 2 for the repeated BIRDS for the 3 conditions. Recall from STS remained constant across trials in all 3 conditions, and analysis of variance indicated no significant differences among the conditions on Trial 8. Long-term storage clearly increased across trials for all 3 conditions, with no significant dif-

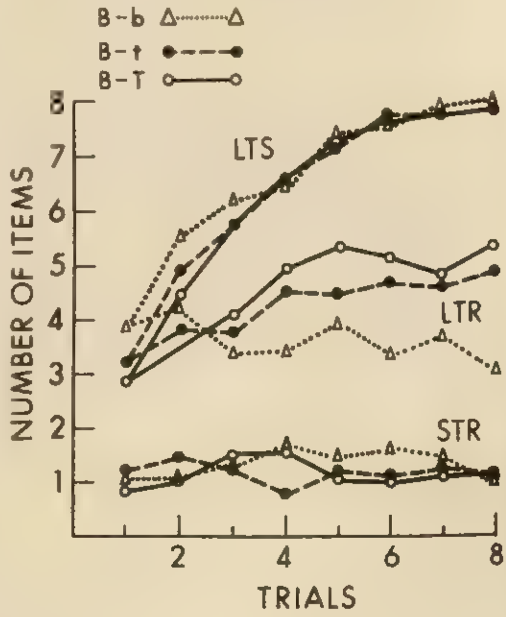


FIGURE 2. Retention in long-term storage (LTS), retrieval from long-term storage (LTR), and recall from short-term storage (STR) for the repeated BIRDS in Conditions B–b, B–t, and B–T.

ferences among the conditions on Trial 8. Retrieval from LTS, however, was significantly different among the conditions on Trial 8, $F(2, 57) = 3.82$, $p < .05$. This indicates that retrieval from LTS accounted for the differences in recall among the conditions shown in the right panel of Figure 1.

DISCUSSION

The results indicate that the context in which the repeated BIRDS were embedded affected their recall. When the rest of the items were from a different category (TREES), changing the TREES to be recalled on each trial resulted in somewhat poorer recall of the repeated BIRDS than when the TREES also were repeated across trials (Condition B–t vs. Condition B–T). Recall of the repeated BIRDS was markedly reduced when the balance of the items, which changed from trial to trial, were from the same category (Condition B–b). In contrast to the other 2 conditions, Condition B–b showed no increase at all across trials in the recall of the repeated BIRDS, even though the repeated BIRDS were presented an equal number of times to each group and in identical serial positions (Figure 1).

The analysis of LTS, retrieval from LTS, and recall from STS (Figure 2) yielded the surprising result that the differences in recall of the repeated BIRDS among the experimental conditions were not due to differences in recall from STS of the repeated BIRDS. We had expected that STS would have been needed for the new items in Conditions B–b and B–t compared to Condition B–T, where *S*s were not confronted with new items on each trial. Even though the new items could have been discriminated for preferential STS in Condition B–t on the basis of category membership, recall from STS of the repeated items was the same as in Condition B–b, where new items could not be distinguished on the basis of category membership. Long-term storage of the repeated BIRDS also was the same in all conditions. Since LTS and recall from STS of the repeated items were identical in all conditions, the differences in recall of these items must be accounted for solely on the basis of differences in retrieval from LTS.

The differences in retrieval from LTS of the repeated BIRDS among conditions reflect the differences in the total composition of the lists presented in the 3 conditions. In Condition B–T, where all of the items were the same from trial to trial, *S*s could learn the items as one coherent list. In Condition B–t, the items could not be treated as part of one coherent list, because half of the list changed on each trial. Vincent curves indicated that as the trials progressed, *S*s in Condition B–t reported an increasingly larger proportion of the new items in the first half of recall and the repeated items in the second half of recall. It appears that *S*s could treat the entire list (B–t) as 2 sublists of repeated (B) and new (t) items. The somewhat poorer retrieval of the repeated BIRDS from LTS in Condition B–t compared to Condition B–T may have been related to the extra processes involved in constructing 2 lists.

In Condition B–b, where the new items on each trial were from the same category as the repeated items, the possibility of constructing 2 lists was impeded by the difficulty in discriminating the repeated from the new items. The Vincent curves also indicated that *S*s in Condition B–b had difficulty in discriminating repeated and new items from the same category because, while they reported more new items in the first half of recall and more of the repeated items in the second half of recall, this was less successful and did occur not until later trials than for *S*s in Condition B–t. The items in Condition B–b could neither be treated as a single coherent list nor as 2 distinct sublists. The *S*s in Condition B–b in effect faced an entirely new list on each trial (even though half of the items were the same on each trial). This is reflected by the failure of the recall of the repeated items to increase across trials.

Taken together, the results of this experiment confirm that the context provided by a list as a whole has a significant effect on the recall of individual items in that list. Recall of the repeated items did not increase over trials when other items from the same category changed from trial to trial. These results also indicate that the differences in recall among the 3 conditions were due to differences in retrieval of the repeated items from LTS, rather than to differences in storage and retention of the repeated items in LTS. Retrieval from LTS appears to depend on the properties of the total list within which the items are presented.

REFERENCES

BOWER, G. H. A selective review of organizational factors in memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

BUSCHKE, H. Spontaneous remembering after recall failure. *Science*, in press.

KINTSCH, W. *Learning, memory and conceptual processes.* New York: Wiley, 1970.

TULVING, E., & COLOTLA, V. A. Free recall of trilingual lists. *Cognitive Psychology*, 1970, 1, 86–98.

(Received April 12, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 729–732

THE EFFECT OF NUMBER OF STIMULI ON THE STEADY-STATE GENERALIZATION GRADIENT:

A TEST OF THE SUMMATION HYPOTHESIS [1]

JOHN C. MALONE, JR.[2]

University of Tennessee, Knoxville

The interacting-gradient model for discrimination learning predicts changes in the slope of the generalization gradient as a function of the number of stimuli comprising it. Attempts to verify this prediction, using the method of extinction testing after reinforced pretraining, have proven inconclusive. In this experiment, three pigeons produced steady-state generalization gradients over a prolonged training period. The number of stimuli forming the gradient was varied, while the temporal appearance of stimuli and the pattern of reinforcement availability remained constant. Variations in the number of stimuli present did not affect the slope of the gradient.

An intuitively reasonable way to view discrimination learning is in terms of excitatory and inhibitory response tendencies; such a position was taken by Pavlov (1927) and was adopted in various forms by others. Probably the clearest formulation of this view was that proposed by Spence (e.g., 1937). According to this interpretation, discrimination learning is a secondary process, dependent upon the formation and interaction of generalization gradients of excitation and inhibition. During the formation of a discrimination between two stimuli on the same dimension (e.g., size, hue), a gradient of excitatory response tendency is presumed generated around S+, as the result of reinforcement (e.g., food) received for responding in its presence, while lack of reinforcement for responding in S− results in an inhibitory gradient around it. The height of these gradients is presumed to decrease with distance from S+ and S−. A simple algebraic summation of such gradients determines the tendency to respond to S+, S−, and to other stimuli which lie within the range of generalization of either.

Despite a good deal of criticism, this model has remained the most popular general account for discrimination learning. In theory, it can account for the learning of a discrimination, for transposition and transposition reversal, and for most of the features of the shifted peak of the postdiscrimination generalization gradient (Hearst, 1968; Hilgard & Bower, 1966). The model has been limited in usefulness, however, due to difficulty in empirically verifying its predictions.

These difficulties must be due in large part to the extreme variability of empirical generalization gradients obtained with present methods, i.e., in the manner popularized by Guttman and Kalish (1956). Animals are usually trained in the presence of one or more stimuli, under conditions of intermittent reinforcement, so that resistance to extinction is developed. Transfer tests are then carried out in extinction; reinforcement is discontinued and the animal is presented a number of stimuli varying along some dimension of the training stimulus.

Two characteristics of empirical gradients which seem essential if they are to be used to explain the phenomena of discrimination learning are stability and recoverability. Gradients obtained in extinction are variable, both among and within subjects (cf. Blough, 1965). Such gradients are not recoverable if testing is repeated after additional single-stimulus training (Zeiler, 1969). The form of these gradients is influenced by a multitude of factors, such as the schedule of reinforcement used during training (Hearst, Koresko, & Poppen, 1964) and the length of the transfer test in extinction (e.g., Guttman & Kalish, 1956).

Several investigators (e.g., Blough, 1965; Zeiler, 1969) have suggested that gradients be obtained in the steady state, i.e., with reinforcement maintained in S+ and with presentation of S+ and "test" stimuli over many sessions. There are several advantages in following this procedure. Steady-state gradients reach quite stable asymptotes (e.g., Malone & Staddon, 1973); they are relatively independent (at asymptote) of the reinforcement schedule in force in S+ (Zeiler, 1969) and, perhaps most important, they are recoverable after interpolated training conditions (Zeiler, 1969).

The interacting-gradient model predicts summation of excitatory and inhibitory response tendencies which should affect the slope of the generalization gradient obtained during extinction testing, as well as during steady-state training. Excitatory gradients summating around reinforced stimuli should produce increased overall responding and a shallower gradient. The addition of nonreinforced stimuli should act to lower overall responding and to increase the slope of the gradient due to the addition

[1] This research was supported by a U.S. Public Health Service Predoctoral Fellowship and by grants from the National Science Foundation to Duke University, J. E. R. Staddon, Principal Investigator, and by funds from a Biomedical Sciences Support Grant to the University of Tennessee, Knoxville.

I thank J. E. R. Staddon, N. Guttman, and I. T. Diamond for their criticisms of an earlier version of this manuscript and S. Forgey and Martha L. Guest for assistance with data analysis.

[2] Requests for reprints should be sent to John C. Malone, Jr., Psychology Department, University of Tennessee, Knoxville, Tennessee 37916.

of inhibitory gradients around these stimuli. Evidence for summation of this sort is equivocal; Kalish and Guttman (e.g., 1957), using the method of extinction testing after reinforced pretraining, did not find convincing evidence for summation of excitatory gradients. Others, using the same method, have found evidence for excitatory-inhibitory summation following pretraining with reinforced and nonreinforced stimuli; the gradient obtained in extinction was shifted toward S+, as would be expected if an inhibitory gradient around S+ is subtracted from an excitatory gradient centered around S− (e.g., Hearst, 1968).

Changes in the number of test stimuli presented during extinction should similarly affect gradient form, due to the summation of inhibitory gradients around these stimuli; such effects have been reported (e.g., Marsh, 1967), but in other cases no effect appears (e.g., Guttman & Kalish, 1956). Others, using steady-state procedures, have varied the number of stimuli comprising the gradient (e.g., Pierrel & Sherman, 1962). However, these studies were not attempts to definitively test the interacting-gradient model; in addition, their procedures did not ensure that factors other than the number of stimuli comprising the gradient remained constant.

The interacting-gradient model is influential enough to warrant testing using the best means available; this experiment varies the number of stimuli comprising a steady-state gradient, while holding other factors constant.

Method. Three white Carneaux pigeons were maintained at 80% of their free-feeding weights throughout the experiment. One bird (119) was experimentally naive; the other two birds (48 and 72) were experienced with multiple fixed-interval schedules (in which key colors served as discriminative stimuli).

The experimental chamber was a 33 × 33 × 33 cm. Plexiglas box with a 1.9-cm. translucent key mounted on one wall. A line orientation produced by a Tektronix (Model 503) oscilloscope was focused onto the key by a small lens (focal length 1 in.) mounted between the oscilloscope face and the key. The image on the key was a line orientation approximately 1.3 × .5 cm., always presented on a blank background. A 24-v. battery connected to the horizontal input of the oscilloscope served to eradicate the trace left on the oscilloscope face between stimulus changes and during reinforcement. A 25-w. red lamp was mounted approximately 8 cm. behind the response key; when lit, the lamp transilluminated the key, providing a red stimulus which could be presented instead of the line-orientation stimuli provided by the oscilloscope. The animal chamber and ⅔ of the oscilloscope were enclosed by a larger box (51 × 51 × 122 cm.) comprised of hollow 5-cm. thick walls filled with soundproofing materials. Reinforcement (3-sec. access to mixed grain) was provided by a food magazine; the 5-cm.[2] aperture was located 5 cm. below the key. The key display was extinguished during reinforcement but the 15-w. fluorescent houselight remained lit. White noise and a ventilation fan masked extraneous noises; programming and recording equip[illegible]ent were located in an adjacent room.

A daily session comprised 90 1-m[illegible]timulus presentations; the sequence of presen[illegible]ns consisted of 10 blocks of nine present[illegible] each. Each block included three presentati[illegible] of S+ (90°; vertical line) and one presentation [illegible]ach of six other orientations (66°, 54°, 42°, 30[illegible]°, and 6° from the horizontal), randomized w[illegible]locks. During S+ (90°) a variable-interv[illegible] 1-min. reinforcement schedule was in effect; VI [illegible] was in effect at all other times. The selectio[illegible]nter-reinforcement intervals for both schedules [illegible]based on the progression suggested by Fleshler [illegible] Hoffman (1962); this was done to ensure steady [illegible]onse rates, by making reinforcement availability [illegible] unpredictable as possible.

The sequence of stimulus presentations was fixed throughout the experiment; although seven orientation stimuli were included in the sequence, a maximum of four appeared during any one cond[illegible]on. Orientations which were not scheduled to appear during a given condition were replaced by presentations of a red key light; VI 3-min. remained in effect during this stimulus. In this way, the temporal appearance of stimuli and the pattern of reinforcement availability during the session remained constant over conditions; *only* the presence or absence of orientation stimuli was varied.

To facilitate control by stimuli on the key, "no-stimulus" periods were inserted at the end of odd-numbered blocks of stimuli. During these periods no stimulus was present on the key and reinforcement was never obtained. The length of the period was determined by a timer which reset with each key peck, prolonging the period. The timer was initially set at 5 sec.; with improved performance (absence of key pecking), the setting was increased to 20 sec.

Each bird received preliminary magazine and key training in the presence of S+ (90°). This was followed by a period of initial training consisting of 11 sessions of 30 1-min. presentations of S+, reinforced according to VI 30 sec., alternated with 30 "no-stimulus" periods.

The experiment comprised the following three conditions: (*a*) S + (90°) vs. 6°; (*b*) S + (90°) vs. 30°, 6°; and (*c*) S + (90°) vs. 54°, 30°, 6°. Each bird was trained on each condition and returned to the original condition, as a test for recoverability. The order of conditions and number of sessions for each bird appear in Table 1.

Results. Within a week after the onset of training, virtually no responding occurred during "no-stimulus" periods. Extended training was carried out in order to assess any changes in gradient form due to length of training; little change occurred after the first few sessions.

There was no change in gradient form as a function of the number of orientation stimuli present. Table 2 shows average response rates per minute and standard deviations during the last four days of each condition for individual birds. Gradients were extremely stable over these periods, as they

TABLE 1

Sequence of Conditions and Number of Sessions of Training for Each Bird

Bird 72		Bird 48		Bird 119	
Condition	Sessions	Condition	Sessions	Condition	Sessions
a	31	*c*	36	*a*	38
b	22	*a*	37	*b*	27
c	16	*b*	5	*c*	11
a	15	*c*	5	*a*	5

were during the entire period of training. Bird 72's gradients were virtually indistinguishable over 69 sessions (Conditions *a*, *b*, and *c*), in spite of changes in the number of orientation stimuli present. Similarly, gradients for Bird 48 varied little across the first three conditions. However, attempts to recover the performance shown in the original condition led to an increase in absolute response rates in all stimuli for Bird 72, and an increase in 54° for Bird 48.

Bird 119 responded at extremely high rates throughout the experiment. During the first two conditions (65 sessions), 119's gradients remained approximately flat (Table 2), reflecting a lack of control by orientation stimuli. During Condition *c*, the gradient steepened, an effect which increased with the shift back to the initial Condition *a*. During a subsequent experiment, this animal showed a similar insensitivity to line-orientation stimuli; thus, the possibility that the lack of previous discrimination experience was responsible for his performance is unlikely. His data will not be discussed further.

The effect of alterations in the number of orientation stimuli on the slope of the gradients appears in Figure 1. The curves in this figure represent linear least squares approximations of the data in Table 1. Gradient slope remained virtually unchanged over conditions.

Discussion. The interacting-gradient model predicts that the slope of the steady-state generalization gradient will change with variation in the number of stimuli comprising it. However, the model may be interpreted to predict two outcomes for the present experiment.

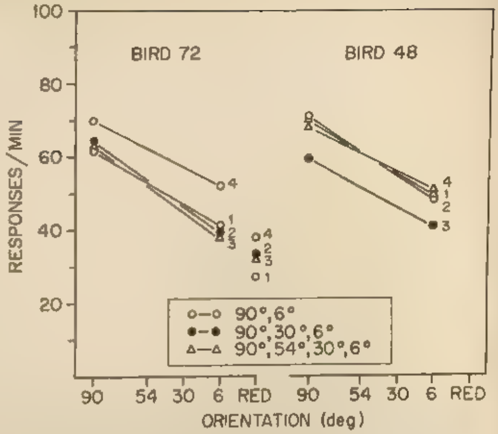


Figure 1. Least squares approximations of data averaged over the last four days of each condition for Birds 72 and 48 (see Table 1). (Gradients are labeled according to the order in which conditions appeared in training for each bird, rather than according to the number of stimuli appearing in that condition.)

Since responding was reinforced in *all* stimuli, though stimuli other than S+ were reinforced at a lower rate, the model would predict that excitatory gradients would be generated around all stimuli. Response rate to a particular stimulus depends upon the excitatory gradient associated with it, as well as upon summation of this gradient with gradients centered around other stimuli on the same dimension (cf. Spence, 1937). Thus, when the number of orientation stimuli forming the gradient is increased, the gradient should *flatten* and overall responding in all stimuli should *increase*, due to the summation of additional excitatory gradients.

An alternative prediction comes from Guttman's (1959) finding that a stimulus associated with

TABLE 2

Mean Response Rates/Minute during the Last Four Days of Each Condition

Condition	90°		54°		30°		6°		Red	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Bird 72										
a	62.0	4.5					41.1	3.7	27.3	2.4
b	65.8	4.7			41.1	4.8	42.9	1.3	33.4	2.0
c	64.4	4.3	52.3	4.9	40.9	2.5	40.5	3.1	32.4	2.6
a	70.1	6.9					52.0	4.8	37.8	3.5
Bird 48										
c	69.8	5.0	64.6	7.9	51.8	5.4	51.8	5.7	30.9	1.7
a	70.6	4.2					48.5	5.9	41.4	1.9
b	58.3	3.4			50.6	5.1	37.7	3.3	39.8	4.4
c	62.0	3.6	70.5	4.9	57.3	4.4	46.4	6.9	36.4	2.1
Bird 119										
a	160.7	4.8					141.4	14.3	96.7	7.7
b	153.6	17.1			146.4	10.2	139.6	12.8	77.5	4.7
c	152.6	6.4	151.4	14.6	127.3	15.7	112.5	14.9	82.4	10.6
a	161.1	6.5					78.1	6.0	85.4	7.5

relatively less frequent reinforcement can become "functionally" inhibitory. Thus, the gradients centered around orientation stimuli other than S+ might be considered inhibitory, rather than excitatory. The summation of such gradients with an excitatory gradient centered around S+ predicts that an increase in the number of orientation stimuli should cause the gradient to *steepen* and response rates in all stimuli to *decrease*.

Neither of these effects occurred; an increase or decrease in the number of orientation stimuli produced no discernible change in gradient slope or in absolute response rates. The only major change in gradient slope during the experiment appeared in the data of Bird 119; this change cannot be attributed to changes in the number of stimuli forming the gradient.

The procedure used here seems optimal for testing the interacting-gradient model; predictions made by either interpretation of the model regarding changes in gradient slope as a function of variations in the number of stimuli present were not confirmed.

REFERENCES

BLOUGH, D. S. Definition and measurement in generalization research. In D. I. Mostovsky (Ed.), *Stimulus generalization*. Stanford: Stanford University Press, 1965.

FLESHLER, M., & HOFFMAN, H. S. A progression for generating variable interval schedules. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 529-530.

GUTTMAN, N. Generalization gradients around stimuli associated with different reinforcement schedules. *Journal of Experimental Psychology*, 1959, **58**, 335-340.

GUTTMAN, N., & KALISH, H. I. Discrimination and stimulus generalization. *Journal of Experimental Psychology*, 1956, **51**, 79-88.

HEARST, E. Discrimination learning as the summation of excitation and inhibition. *Science*, 1968, **162**, 1303-1306.

HEARST, E., KORESKO, M. B., & POPPEN, R. Stimulus generalization and the response reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, 1964, **7**, 369-380.

HILGARD, E. R., & BOWER, G. H. *Theories of learning* (3rd ed.). New York: Appleton-Century-Crofts, 1966.

KALISH, H. I., & GUTTMAN, N. Stimulus generalization after equal training on two stimuli. *Journal of Experimental Psychology*, 1957, **53**, 139-144.

MALONE, J. C., JR., & STADDON, J. E. R. Contrast effects in maintained generalization gradients. *Journal of the Experimental Analysis of Behavior*, 1973, **19**, 167-179.

MARSH, G. Inverse relationship between discriminability and stimulus generalization as a function of the number of test stimuli. *Journal of Comparative and Physiological Psychology*, 1967, **64**, 284-289.

PAVLOV, I. P. *Conditioned reflexes*. (Trans. by G. V. Anrep) New York: Dover, 1927.

PIERREL, R., & SHERMAN, J. G. Generalization and discrimination as a function of the $S^D - S^\Delta$ intensity difference. *Journal of the Experimental Analysis of Behavior*, 1962, **5**, 67-71.

SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, 1937, **44**, 430-444.

ZEILER, M. D. Repeated measures of reinforcement effects on gradients of stimulus control. *Journal of the Experimental Analysis of Behavior*, 1969, **12**, 451-461.

(Received April 17, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 732-734

OVERHABITUATION AND SPONTANEOUS RECOVERY OF THE GALVANIC SKIN RESPONSE[1]

JAMES P. JAMES[2]
Fort Hays Kansas State College

KEN R. DANIELS
Brandon University

BRIAN HANSON
University of Manitoba, Winnipeg, Manitoba

The time course of recovery from 2 levels of habituation was examined. In a 2 × 4 completely randomized factorial design, with 24 *S*s per cell, the criteria of habituation were 2 or 5 successive trials without a response, and the recovery intervals were 1, 2, 4, or 6 min. Recovery of the orienting response was an increasing negatively accelerated function of the recovery interval. The 2 criteria of habituation did not produce significantly different curves of recovery. When both criterion groups were combined at each recovery interval, the median percentages of recovery for the 1-, 2-, 4-, and 6-min. intervals were 17%, 55%, 81%, and 84%, respectively.

The term *spontaneous recovery* refers to the response increment that occurs after a "rest" period which follows either extinction of a learned response or habituation of an orienting response (OR). Despite the fact that spontaneous recovery of a habituated OR is considered a major parametric feature of habituation (Thompson & Spencer, 1966), little systematic research with human *S*s has been reported.

[1] The results of this research were presented at the meeting of the Rocky Mountain Psychological Association, Albuquerque, May 1972.

[2] Requests for reprints should be sent to James P. James, who is now at the Department of Psychology, Western Illinois University, Macomb, Illinois 61455.

The purpose of the present study was to investigate the spontaneous recovery of the galvanic skin response (GSR) component of the OR. Specifically, an attempt was made to examine the time course of recovery from 2 levels of habituation.

Method. The *S*s were 192 male and female students enrolled in a general psychology course. Participation in the experiment partially fulfilled a course requirement.

The apparatus consisted of a Hunter GSR amplifier, modified to detect resistances from 0 to 150,000 ohms. Fels 20-mm.-diam. zinc electrode cups, filled with Fels electrode jelly, were used. Tones from a Grason–Stadler 950-D generator were presented through Roberts headphones. Stimulus durations were timed by a Hunter interval timer. The intertrial interval (ITI) was timed by *E* with a Standard timer, and the recovery interval was timed with a stopwatch.

The electrodes were attached to the palm and back of *S*'s right hand. The *S* was told that he/she would hear several short tones through the headphone and that he/she was to relax and remain as still as possible. The *E* next assigned each *S* randomly to 1 of 8 groups based on a completely randomized factorial arrangement of 2 habituation levels and 4 recovery intervals. The 2 habituation levels were 2 and 5 successive trials without a response, as defined later. The 4 recovery intervals were 1, 2, 4, and 6 min. following the final trial of habituation. These intervals were defined as the time between the onsets of the final habituation stimulus and the test stimulus.

After the baseline resistance became relatively stable, 1,000-Hz. tones of 2-sec. duration were presented at random intervals of 30, 40, or 50 sec. ($M = 40$ sec.). The tonal intensity was 62 db, measured at the headphone. Stimulus presentations continued until the appropriate criterion of habituation was met. On the test trial, the measure of recovery was the GSR magnitude elicited by the same standard stimulus.

A response was defined as a decrease in skin resistance of at least 200 ohms occurring 1–5 sec. after stimulus onset. These defined changes in skin resistance were transformed to log conductance change scores. A base-level resistance score was the skin resistance in ohms recorded at stimulus onset.

Results. On the first trial of habituation, the 8 groups did not differ in either mean GSR magnitude ($F < 1$) or base-level resistance ($F < 1$). On the final trial of habituation, the groups did not differ in base-level resistance, F (7, 184) = 1.16, $p > .20$.

Inter-*S* variability in rate of habituation was high. For the 2- and 5-trial groups, the ranges of trials required to meet the criterion of habituation were 3–30 and 6–27, respectively. The 2-trial groups averaged 7.29 trials to meet criterion; the 5-trial groups required a mean of 10.66 trials. Trials to criterion of habituation did not differ significantly within the 2-trial ($F < 1$) or the 5-trial, F (3, 92) = 1.58, $p > .10$, groups.

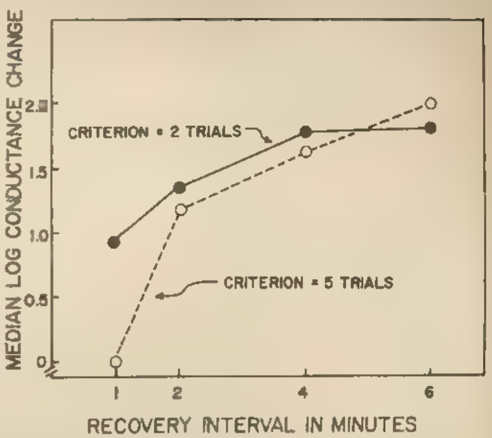


FIGURE 1. Median galvanic skin response magnitude across the recovery interval as a function of 2 criteria of habituation.

In Figure 1, median GSR magnitudes for the 2 criteria of habituation are plotted across the recovery intervals. Each data point represents a single group's median GSR magnitude. Because of several zero scores, with the resulting deviance from normality, nonparametric statistics were used for data analysis. By Kruskal-Wallis tests (corrected for ties), both 2-trial, H (3) = 9.88, $p < .02$, and 5-trial, H (3) = 29.31, $p < .001$, groups exhibited significant increases in response magnitude across the recovery interval. Although differential rates of recovery are suggested in Figure 1 by the apparent difference between the groups at the 1-min. interval, a Mann-Whitney U test did not yield an acceptable level of significance ($Z = 1.28$, $p = .10$, 1-tailed test).

The percentage recovery was calculated for each S by dividing the GSR magnitude of the recovery test trial by the equivalent score for the first habituation trial. When the 2 criterion groups were then combined at each recovery interval, the median percentages of recovery for the 1-, 2-, 4-, and 6-min. intervals were 17%, 55%, 81%, and 84%, respectively.

Discussion. The results suggest that the GSR component of the OR recovers a considerable portion of its prehabituation response magnitude. Fast recovery to the 2-min. posthabituation point was followed by slower recovery to 6 min. This increasing negatively accelerated curve also describes recovery from habituation of the withdrawal and hooking response in earthworms (Gardner, 1968), the head-shake response in rats (Leibrecht & Askew, 1969), and a variety of responses in several other species of vertebrates and invertebrates (cf. Hinde, 1966).

The 2 criteria of habituation used here did not produce significantly different curves of OR recovery, a finding consistent with those of Gardner (1968). However, it is clear that the present study and that of Gardner can be compared only in general terms, because of the differences in response systems and species investigated.

The present results raise the question of how many habituation trials beyond a certain point, a zero response being an arbitrary point, are necessary to produce differential recovery of the OR. If the ITI is thought of as one of the stimulus parameters, any lengthening of the ITI, such as occurs during the recovery interval, would represent a change in stimulus that may serve to evoke a generalized OR. Sokolov's (1960) neuronal model theory of habituation predicts that up to a point, the greater the number of habituation trials, the more completely established the model. A change in any stimulus parameter would then result in a greater mismatch between stimulus and model. Since overhabituation necessarily involves additional habituation trials, the increased stimulus novelty produced by the recovery interval may evoke a response increment that would tend to counteract the decremental effects of these additional habituation trials. This speculation predicts steeper recovery curves is a function [illegible] habituation trials, but may be limited to [illegible] where relatively few habituation trials are [illegible]tered. A steeper recovery curve for the group [illegible]eived the greater number of habituation trials [illegible]ested by the present data.

REFERENCES

GARDNER, L. L. Retention and overhab[illegible] [illegible]lual-component response in *Lumbricus* [illegible] *of Comparative and Physiological Psychology*, [illegible] [illegible]318.

HINDE, R. A. *Animal behaviour: A synthesis* [illegible] *and comparative psychology*. New York: McGraw-[illegible]

LEIBRECHT, B. C., & ASKEW, H. R. Habituat[illegible] [illegible]ead-shake response in the rat: Recovery, transfer, [illegible] [illegible]nges in topography. *Journal of Comparative and* [illegible] *Psychology*, 1969, **69**, 699–708.

SOKOLOV, E. N. Neuronal models and the orient[illegible] [illegible]flex. In M. A. B. Brazier (Ed.), *The central nervous* [illegible] *and behavior*. New York: Josiah Macy, Jr. Foundat[illegible] [illegible]960.

THOMPSON, R. F., & SPENCER, W. A. Habituation: [illegible] [illegible]del phenomenon for the study of neuronal substrates of [illegible]ior. *Psychological Review*, 1966, **73**, 16–43.

(Received April 21, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 734–736

ASSOCIATIVE CONFUSIONS IN MENTAL ARITHMETIC [1]

JOHN H. WINKELMAN [2] AND JANET SCHMIDT

University of Oregon

The effects of associative interference between single-digit addition and multiplication problems were investigated. Associative confusions (e.g., 4 + 5 = 20) were about 50 msec. slower than comparable nonassociative confusions (e.g., 4 + 5 = 15), and produced many more errors. The result indicates the presence of associative processes in simple mental arithmetic and casts doubt on the counting theory of addition.

Previous work on single-digit addition problems has focused on reaction times for verifying sums as a function of the size of the addends (Parkman, 1971; Parkman & Groen, 1971). Parkman and Groen found a strong linear relationship between the size of the minimum of the two addends and the reaction time required for verifying the sum. They accounted for this finding by positing an unconscious counting process: first, the maximum of the addends is determined, and then it is incremented by ones for a number of times equal to the minimum addend. Since each increment presumably uses up the same amount of time, the decision time should be a linear function of the minimum of the two addends.

The counting model is not the only model consistent with the linear increase of reaction time with minimum addend. It is possible that addition is accomplished by an association between a digit pair (the addends) and the sum. As the size of the minimum digit increases, the strength of association between the digit pair and the sum decreases (because larger sums are practiced less often), and hence the verification reaction time increases. The associative model would predict a monotone increase of reaction time with minimum digit, rather than a linear increase; but most monotone functions can be fit fairly well by linear functions, so the Parkman and Groen (1971) result would be accounted for.

The present study will not argue that associative processes are the only ones involved in addition and multiplication, but it will attempt to establish that associative processes do occur in addition and multiplication. It is proposed that there exist associations between pairs of digits and both their sums and products; for example, the digit pair (3, 3) would have associations with both 6 and 9. The obvious prediction is that stimuli of the type 3 + 3 = 9 and 3 × 3 = 6 will produce a tendency to respond *yes* because of associative interference, and hence that stimuli of this type will produce longer

[1] This research was supported by the following grants: National Institute of Education Grant 3-0097; National Institute of Mental Health Grant MH-17958-04; and Advanced Research Projects Agency Grant 50-262-5304.

[2] Requests for reprints should be sent to John Winkelman, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

reaction times and more errors than comparable stimuli not exhibiting interfering associations.

Method. Six University of Oregon students served as *Ss*, each for 1 hr. on three separate days.

The stimuli were constructed from the pairs (3, 3), (4, 3), (3, 5), (4, 5), and (5, 5). Each pair could serve as part of an addition or a multiplication stimulus. The stimulus answer (the right-hand side of the equation) was one of the following four types: (*a*) associative confusion items (e.g., $3 + 5 = 15$; $3 \times 5 = 8$); (*b*) nonassociative confusion items, chosen from the set of possible answers to the other operation (e.g., $3 + 5 = 25$, because 25 is in the set of answers to multiplication problems); (*c*) confusion items, chosen from the set of possible answers for the actual operation (e.g., $3 + 5 = 10$, because 10 is the answer to one of the addition problems); or (*d*) the correct answer (e.g., $3 + 3 = 6$). The complete set of stimuli is given in Table 1. Each number which occurs on the right hand side of an equation is the correct answer for some problem in the set.

Subjects were run under three modes of presentation. In the blocked mode, *S*s received 30 addition problems, then a "ready" signal, and then 30 multiplication problems; the 60 problems together constituted one block in the design. In the triad mode, *S*s received three addition problems, then three multiplication problems, then three addition problems, etc., for 60 trials; the 60 problems constituted a block. In the random mode, *S*s received 30 addition problems and 30 multiplication problems in random order, so that *S*s did not know whether to expect an addition or a multiplication problem. In all three modes, the stimuli were selected by a random process, without replacement.

Within each block, half of the stimuli were correct and half were incorrect. Each stimulus of the three confusion types occurred exactly once per block, and each correct stimulus occurred exactly three times per block. Subjects were given a practice block, which was not recorded, and then eight blocks of trials under a single mode within a day. Mode of presentation was balanced against days in a Latin square design, so that each mode occurred an equal number of times on the first, second, and third days, and so that each *S* received each mode.

A visual warning signal preceded each test stimulus by 500 msec. Subjects were required to push the left of two keys if the stimulus was correct, and the right key if it was not correct; the test stimulus remained on until the *S* made a response. Feedback was given after each response.

The stimuli were exposed on a cathode ray tube, under control of the PDP-15 computer of the University of Oregon cognitive laboratory.

Results. Only data from the three confusion types were analyzed; the stimuli for which the correct answer was *yes* did not bear on the hypothesis.

The experimental design is given by SAMCB, where S = Subjects, A = Operation (addition or multiplication), M = Mode of presentation (blocked, triad, or random), C = Confusion type (associative,

TABLE 1

STIMULI USED IN THE EXPERIMENT

Operation	Associative confusions	Nonassociative confusions	Target confusions	Correct stimuli
$3 + 3 =$	9	12	7	6
$4 + 3 =$	12	9	6	7
$3 + 5 =$	15	25	10	8
$4 + 5 =$	20	15	8	9
$5 + 5 =$	25	20	9	10
$3 \times 3 =$	6	7	12	9
$4 \times 3 =$	7	6	9	12
$3 \times 5 =$	8	10	25	15
$4 \times 5 =$	9	8	15	20
$5 \times 5 =$	10	9	20	25

nonassociative, or target), and B = Block (eight levels).

For the reaction time data, medians were taken over the five stimuli occurring within a block, confusion type, and level of A. Thus, each block yielded six medians. An analysis of variance was performed on these medians.

There was a significant effect of Confusion type, $F\ (2, 10) = 12.96$, $p < .005$, and of Block, $F\ (7, 35) = 2.37$, $p < .05$. Two interactions were significant: the Operation $\times$ Confusion type interaction, $F\ (2, 10) = 18.27$, $p < .001$, and the Mode of Presentation $\times$ Confusion type interaction, $F\ (4, 20) = 4.20$, $p < .05$.

The Scheffé procedure (Meyers, 1966, p. 333) was used to test whether the differences between particular confusion types were significant. The comparison of chief interest was between associative and nonassociative confusions. An experiment-wide error rate of .10 was used, giving an F critical of 5.48. The associative–nonassociative comparison gave an F of 25.90; the associative–target comparison gave an F of 6.90; and the nonassociative–target comparison gave an F of 6.90.

Associative confusions are consistently about 50 msec. slower than nonassociative confusions, across both addition and multiplication, and in all three modes of presentation. Every *S* showed the effect. The error data show 221 errors for associative confusions, as opposed to 46 errors for the nonassociative confusions.

The main effect of Blocks was due to a fairly small but consistent decrease in reaction times over blocks, a practice effect. The difference between associative and nonassociative confusions did not show any consistent decrease over blocks.

The absence of a main effect for Mode of Presentation is surprising: one would have expected the Blocked mode to be faster than the Random mode. The cell means show a tendency in this direction, but the differences are not significant (due to a large Subject $\times$ Mode of Presentation error term, which includes variation arising from the differing orders in which *S*s received the Blocked, Triad, and Random conditions). The error data show the Blocked and Random modes to be approximately equal (mean errors = 3.06 and 3.17, respectively),

TABLE 2

TABLE OF MEAN REACTION TIMES (IN MSEC.)

Item	Confusion type			
	Associative	Nonassociative	Target	[illegible]
Operation				
Addition	749.5(5.2)	695.5(5.2)	691.5(5.2)	713.[illegible]
Multiplication	742.1(5.2)	685.3(5.2)	739.9(5.2)	722.4
M	745.5(7.7)	690.4(7.7)	712.2(7.7)	
Mode of presentation				
Blocked	709.1(4.3)	645.8(4.3)	698.2(4.3)	687[illegible]
Triad	766.7(4.3)	709.8(4.3)	732.3(4.3)	736.[illegible]
Random	761.6(4.3)	706.6(4.3)	721.1(4.3)	729.8
M	745.5(7.7)	690.4(7.7)	712.2(7.7)	

Note. Estimates of standard error derived from [illegible]

with the Triad mode producing somewhat more errors than either (4.03).

The interaction of Operation with Confusion type results solely from the different behavior of Target confusions under addition and multiplication. The weak interaction between Mode of Presentation and Confusion type is also due to the behavior of Target confusions under the three modes of presentation.

Discussion. The main result of interest is that associative confusions produce slower reaction times and more errors than nonassociative confusions. This result cannot be attributed solely to the readings by *S*s, because the effect occurs under all three modes of presentation: in both the blocked and triad modes the *S*s knew which operation to expect in the next problem.

TABLE 3

MEAN ERRORS PER SUBJECT

Item	Statistic	
	M	SE_m
Mode of presentation		
Blocked	3.06	.56
Triad	4.03	.59
Random	3.17	.60
Confusion type		
Associative	6.03	.71
Nonassociative	1.28	.24
Target	2.94	.38

The result forces us to reject an explanation of addition or multiplication based solely on internal computations since such a model would predict no difference between associative and nonassociative confusions. The result strongly suggests that the processes of addition and multiplication contain associative components, and is consistent with the theory that only associative processes are involved in addition and multiplication.

REFERENCES

MEYERS, J. L. *Fundamentals of experimental design.* Boston: Allyn and Bacon, 1966.

PARKMAN, J. M. Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology,* 1971, **91**, 191-205.

PARKMAN, J. M., & GROEN, G. J. Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology,* 1971, **89**, 335-342.

(Received April 23, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 737–739

CUE EFFECTIVENESS IN CUED RECALL[1]

MARION QUINN LEWIS[2]

State University of New York College at Plattsburgh

Two experiments investigated effects of cues for categorized lists with both blocked and unblocked category presentation. Category labels and extralist category items were effective recall cues for both blocked and unblocked category lists. Cuing increased recall of categories but did not affect words recalled per category.

With categorized lists, providing the category labels or items from the categories as cues has been found to facilitate recall if category members were presented contiguously (blocked), but cuing has not been found to increase recall if category items were mixed (unblocked) in the lists (Dong, 1972; Lewis, 1971; Slamecka, 1968, 1972; Tulving & Pearlstone, 1966).

Facilitation of recall as a function of cuing has been interpreted as evidence that more information is available in memory than is immediately accessible for recall. Tulving and Pearlstone (1966) interpreted recall of categorized lists as a product of two independent processes, recall of "higher order memory units" (the categories) and the recall of items within the higher order units (category items). Category recall is seen as depending on the number of categories and retrieval cues, but recall of items within a category is apparently related only to whether the category itself is recalled and not to the experimental cues. Cuing increased the number of categories recalled, but did not affect the number of items recalled per category.

The failure of cuing to facilitate recall with unblocked category presentation is puzzling because it is not apparent that the two-level process described by Tulving and Pearlstone (1966) should be specific to blocked lists.

It should be noted that most investigators reporting no cued facilitation with unblocked lists used items from the list as cues (Lewis, 1971; Slamecka, 1968). List-item cues do not appear to be without effect, as Lewis (1971) reported such cues increased recall of blocked lists. However, when list items are used as cues, scoring of both cued and noncued conditions has been on the basis of "critical" words. That is, cue words are chosen before the list is presented, and the other list items are regarded as critical or target items. Thus, to the extent that *S*s recall noncritical items which are not scored, total recall may be underestimated. It may be possible to devise scoring procedures or experimental manipulations which are unbiased in this respect. For example, this bias would not occur if cues were category labels (e.g., Tulving & Pearlstone, 1966) or other extralist items.

In the present investigation of cuing effects with blocked and unblocked category presentation, two types of cues were used: category labels and category members which were not presented as list items.

Experiment I: Category Label Cues

Method. Eight categorized lists were prepared. Each 42-word list consisted of six Battig and Montague (1969) categories, with seven words per category. Category items were the words ranked 2–8 in the association norms, excluding obvious multiple-category items (for example, orange as color and fruit). Categories were assigned to lists at random. Both a blocked and an unblocked version of each list were prepared. For the blocked version, all category members were presented contiguously; category orders and within-category item orders were randomized. For the unblocked version, category items were mixed throughout the list with the restriction that no more than two items from the same category were presented contiguously.

There were thus four treatments representing lists and cuing conditions: blocked and unblocked list versions, and cued and noncued recall. Each treatment combination was presented twice; four list–treatment combinations and orders were prepared. For Cued conditions, the cue for each category was the category label.

Lists were presented on audio tape at a 2-sec. rate for written free recall. After the last word on each list, the *S*s heard a three-digit number. They wrote that number and subtracted from it for 6 sec., until signaled to begin recall. This procedure was used to equate cued and noncued recall in terms of possible forgetting from short-term memory while examining cues under cued recall (Lewis, 1971; Slamecka, 1968).

When the recall signal was given, the *E* turned over a card. For cued conditions, cues were printed on the card; for noncued recall, the card simply said to begin recall. For any given list, the *S*s did not know whether recall would be cued or noncued until they saw the card. The recall interval was 2 min. The *S*s were tested in small groups of two or three.

The *S*s were 40 students enrolled in psychology classes at the State University of New York College at Plattsburgh. They received extra course credit for participation.

[1] The author thanks C. J. Wallace, Loyola University, and Earl E. Larre, St. Mary's Dominican College, for their cooperation in a preliminary study. An earlier version of this paper was presented at the meeting of the Psychonomic Society, St. Louis, November 1972.

[2] Requests for reprints should be addressed to Marion Quinn Lewis, State University of New York College, Plattsburgh, New York 12901.

TABLE 1

PERCENT TOTAL RECALL, CATEGORY RECALL, AND WORDS PER CATEGORY: EXPERIMENT I (CATEGORY LABEL CUES)

Recall condition	Presentation: Blocked			Presentation: Unblocked		
	Total recall	Categories	Words per category	Total recall	Categories	Words per category
Cued	53.4	95.8	55.7	48.4	97.1	49.3
Noncued	40.6	70.2	57.5	42.7	82.7	51.3

Total recall, number of categories recalled (defined as recall of at least one item from the category) and mean words per recalled category were examined by separate 2 × 2 × 40 (Presentation condition × Cuing condition × *Ss*) analyses of variance. Newman-Keuls comparisons among individual treatment combinations were made.

Results. Total recall, category recall and mean words per recalled category (WPC) are shown by percentages in Table 1.

Total recall was higher for cued recall (CR) than for Noncued recall (NCR), F (1, 39) = 50.48, $p < .005$. For both Blocked and Unblocked presentations, CR was higher than NCR, $p < .01$. The difference between Cuing conditions was larger for Blocked than for Unblocked presentation, F (1, 39) = 10.46, $p < .005$.

More categories were recalled with CR than NCR, F (1, 39) = 120.45, $p < .005$; and category recall was higher for Unblocked than for Blocked presentation, F (1, 39) = 22.22, $p < .005$. The effect was clear in the interaction, F (1, 39) = 10.49, $p < .005$. With CR there was no difference between Blocked and Unblocked presentation, both being nearly perfect; but with NCR, Unblocked category recall was higher than Blocked, $p < .01$.

More words per category (WPC) were recalled with Blocked than with Unblocked presentation, F (1, 39) = 30.50, $p < .005$. No other effects or interactions were significant.

The results are clear. Category label cues facilitate recall for both Blocked and Unblocked lists. The facilitation derives from higher category recall with cues. There is no difference between Cuing conditions in mean words per category.

It is perhaps possible that there is a qualitative difference between category items and category labels as cues. Intuitively, the label seems a more potent cue. In Experiment II, the cue for each category was an item which had not been presented on the list. That is, the cues were extralist members of list categories.

EXPERIMENT II: EXTRALIST CATEGORY ITEM CUES

Method. The lists prepared for Experiment I were used. For Cued conditions, cues were the highest ranking word in each category in the word association norms. The procedure was the same as for Experiment I.

The *Ss* were 26 students enrolled in psychology classes at the State University of New York College at Plattsburgh. They received extra course credit for participation.

Total recall, number of categories recalled, and mean words per recalled category were analyzed by separate 2 × 2 × 26 (Presentation condition × Cuing condition × *Ss*) analyses of variance. Newman-Keuls comparisons among individual treatment combinations were made.

Results. Total recall, category recall and mean words per recalled category (WPC) are shown by percentages in Table 2. Total Cued recall (CR) was higher than Noncued recall (NCR), F (1, 25) = 14.06, $p < .005$; and Blocked presentation was higher than Unblocked, F (1, 25) = 8.77, $p < .01$. The interaction was not significant.

Category recall was higher for CR than for NCR, F (1, 25) = 44.30, $p < .005$. Unblocked category recall was higher than Blocked, F (1, 25) = 7.08, $p < .025$, largely because of higher Unblocked NCR category recall; however, the interaction was not significant.

Mean words per recalled category (WPC) was higher for Blocked than for Unblocked presentation, F (1, 25) = 22.09, $p < .005$. There was no interaction or effect due to Cuing.

Overall, the results replicate those of Experiment I. Cuing facilitates recall by increasing category recall. Mean WPC is not affected.

Discussion. The cued facilitation with both Blocked and Unblocked category presentation using category labels and extralist category items as cues is consistent with Tulving and Pearlstone's (1966) formulation of recall as a process of (*a*) recalling higher order memory units, and (*b*) recalling contents of the higher order units. Cuing increased category recall (higher order units), but did not affect recall of items within a category.

Previous results (Lewis, 1971; Slamecka, 1968) showed no increase in recall for unblocked lists with list-item cues. Present findings indicate consistent facilitation with extralist cues, regardless of presentation condition. This suggests that critical-item scoring procedures used with list-item cues may

TABLE 2

PERCENT TOTAL RECALL, CATEGORY RECALL, AND WORDS PER CATEGORY: EXPERIMENT II (EXTRALIST CATEGORY ITEM CUES)

Recall condition	Presentation: Blocked			Presentation: Unblocked		
	Total recall	Categories	Words per category	Total recall	Categories	Words per category
Cued	50.9	94.9	53.5	45.0	96.2	46.8
Noncued	43.4	76.6	57.1	40.5	83.9	48.1

introduce sizable artifacts, which would account for the apparent discrepancy between previous and present findings.

Slamecka (1972) pointed out that cuing does not increase recall if unaided category recall is nearly perfect. In this context, the smaller differences between cuing conditions observed with Unblocked presentation is understandable. Noncued category recall is substantially higher for Unblocked than for Blocked presentation. There is less opportunity for cuing to add categories.

The higher Unblocked category recall might reflect differential Blocked–Unblocked memory organization. With Blocked presentation, it is most probable that *S*s will use *E*-defined categories to organize information for storage and retrieval. With Unblocked presentation, *S*s may develop somewhat different categories (for example, instead of a single "bird" category, *S* may use two, as "songbirds" and "birds of prey"). Moreover, serial position effects favor higher category recall with Unblocked presentation. Although the distractor task between list presentation and recall tends to reduce the characteristic free-recall recency peak, such a task has little effect on primacy (Glanzer & Cunitz, 1966). The first few words on an Unblocked list are from several categories, but the first few words on a Blocked list are all from the same category.

Taken overall, the results of the present experiments demonstrate that cuing can facilitate recall with both Blocked and Unblocked lists. As such, these findings appear inconsistent with those of Dong (1972), who reported no cued facilitation with category label cues on an unblocked list. Dong presented a given *S* with only one list; in the present experiments *S*s received several lists, both cued and noncued. Lewis (1971) noted that on every measure (total recall, category recall and WPC) differences between cuing conditions were greater for the last list of a series than for the first list. Although *S*s in the present experiments could not anticipate cues for any particular list, it may be speculated that they learned to use cues more effectively after having received cues several times. This speculation is consistent with both Lewis' (1971) findings and Dong's suggestion that cuing effects may reflect strategies and memory control processes.

REFERENCES

Battig, W. F., & Montague, W. F. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category norms. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt. 2).

Dong, T. Cued partial recall of categorized words. *Journal of Experimental Psychology*, 1972, **93**, 123–129.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

Lewis, M. Q. Categorized lists and cued recall. *Journal of Experimental Psychology*, 1971, **87**, 129–131.

Slamecka, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology*, 1968, **76**, 504–513.

Slamecka, N. J. The question of associative growth in the learning of categorized material. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 324–332.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

(Received May 12, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 740–742

RECALL OF ANTONYMS FROM SHORT-TERM MEMORY[1]

LAIRD S. CERMAK[2]

Psychology Service, Boston Veterans Administration Hospital

An experiment was designed to investigate the effect of introducing antonyms on adjacent trials of a short-term memory distractor task. It was discovered that recall of the antonyms increased when the trial on which they occurred was preceded by several trials each containing words that were antonyms of the final trial. However, recall did not increase when only the immediately preceding trial contained the antonyms. An interpretation based on the priming hypothesis is given.

Proactive facilitation (PF) has been used to describe instances in which retrieval of a particular word is facilitated by the fact that the word shares a feature in common with one or more previously presented words. Proactive facilitation has been demonstrated by Cermak (1970), who found, during the course of a distractor experiment, that when the to-be-remembered words in a particular triad are acoustically identical to words presented on the preceding trial (i.e., they are homonyms) their retrieval is facilitated. However, PF does not occur if the relationship between two adjacent trials is based upon the semantic features of the words. For example, synonyms, antonyms, and highly associated words all failed to produce facilitated retrieval of the second member of the relationship.

The failure of synonyms and associates to produce an increase in recall can be explained by the fact that these words also all belong to the same class of material and probably interfere with, rather than facilitate, one another's retrieval. However, interference between antonyms was unexpected since these words come from "opposite" categories and should be differentially encoded. One possible explanation for the antonym result might be that the procedure that was used, one in which only two words shared the critical relationship, may not have been sufficient to produce differential encoding. The usual procedure used to study differential encoding ability presents several instances of a particular relationship before testing for the ability to detect and utilize a differential relationship (Wickens, 1970). Perhaps a subject cannot encode antonyms as belonging to two separate categories unless a differential "relationship" has already been established. The present report investigates this possibility by preceding the introduction of an antonym relationship with either several unrelated word triads (Experiment I) or with several synonymous word triads (Experiment II).

EXPERIMENT I

Method. The *S*s were 48 introductory psychology students at Tufts University who chose to participate in the experiment as partial fulfillment of a course requirement. The *S*s were randomly assigned in order of their appearance to one of the two groups, with 24 in each group.

The short-term memory (STM) distractor technique was used. In this procedure a word triad was projected on a screen and the *S* repeated the words once; the *S* was then distracted from rehearsal until a signal for recall was presented. The distractor task consisted of counting backward by threes from a number flashed on a screen in time with a metronome beating at one beat per second. Four additional trials followed consisting of different word triads but using the same procedure.

The words used in this experiment were chosen in such a way that an antonym existed for each word. Five such word triads were used so that each triad could appear in each of the five positions in the series. In other words, five basic slides were counterbalanced with the fifth trial bearing an antonym relationship to the fourth trial for the subject in the Experimental (Antonym) Group, but no relationship at all for the subject in the Control Group.

The experiment consisted of five trials in which the material presented on Trial 5 bore one of the following relationships to the material presented on Trial 4: Antonym Group—antonymity; Control Group—no relationship. Trials 1–4 bore no relationship to one another for either group. In order to compare the Antonym Group's performance on the fifth trial with the Control Groups', both groups received the same stimulus material on that trial. The antonym or neutral relationship was established by presenting different material on Trial 4. In other words, the Antonym Group received antonyms of Trial 5 on Trial 4, whereas the Control Group received words unrelated to those of Trial 5 on Trial 4. Any differences in recall on Trial 5 could thus be attributed to the effects of the relationship that had been produced previously and not the difficulty level of Trial 5 material.

The *S* was seated facing a screen upon which the material was projected by means of a Kodak Carousel 800 projector. The projector was timed by a Gerbrands tape timer. The only other apparatus

---

[1] This research was supported in part by National Institute of Mental Health Grant MH-21234-01 to the author. Appreciation is extended to Alan Sholk and John Harrison for their help in the collection of data.

[2] Also of the Aphasia Research Unit, Neurology Department, Boston University School of Medicine. Requests for reprints should be sent to Laird S. Cermak, Veterans Administration Hospital, 150 South Huntington Avenue, Boston, Massachusetts 01230.

in the room was a metronome used to pace *S*'s distractor task. After the instructions had been given, *S*s practiced on the counting task until approximation to the metronome was achieved. The timing and sequence of events were then initiated by the tape timer and the time intervals for one trial were as follows: asterisk (ready signal), 2 sec.; trial presentation, 2 sec.; retention interval, 15 sec.; recall interval, 10 sec.; rest interval, 6 sec. Recall was scored as the number of correct words given out loud during the 10-sec. recall interval. Following the end of the fifth trial, *S*s were asked whether they had noticed any similarities or differences among any of the word triads, and their responses were noted at that time.

Results. On Trial 5, the critical trial, the Antonym Group recalled 60% of the words correctly, while the Control Group recalled 57% correctly (Figure 1). There was no significant difference in recall between the group receiving an antonym triad on the previous trial and a group that did not. Furthermore, only two *S*s reported detecting the antonymity relationship. The buildup of proactive inhibition (PI) from Trial 1 to Trial 4 was significant ($p < .01$) for both groups and neither group showed any significant increase from trial 4 recall to trial 5 recall.

EXPERIMENT II

Method. The *S*s were 40 introductory psychology students at Tufts University who chose to participate in the experiment as partial fulfillment of a course requirement. The *S*s were randomly assigned in order of their appearance to one of the two groups with 20 in each group.

The distractor technique was used but, unlike Experiment I in which a relationship existed only between Trials 4 and 5, the present experiment presented word triads on Trials 1 through 4 that were synonymous with each other. In other words, the first word in each triad meant the same as the first word in each of the other triads and the second word meant the same as each other triad's second word, and so on. Trial 5 (the critical trial) contained words that were either antonyms of the preceding four triads, or were synonyms. Thus, the groups were differentiated on the basis of the type of relationship existing between the words presented on Trial 5 with those on Trials 1–4. It must be pointed out that the words appearing on Trial 5 were the same for both groups, and, in fact, were the same as those used in Experiment I. Trials 1–4 were all antonyms of Trial 5 for the Antonym Group, but they were synonymous with one another. For the Synonym Group the same word triad appeared on Trial 5, but all the preceding triads were synonyms of this triad and were synonymous with one another. Two sets of critical Trial 5 word triads were used in the experiment, both appearing under synonym and antonym conditions.

Results. On Trial 5, the Antonym Group recalled significantly more words than the Synonym Group, t (38) = 2.20 ($p < .05$). The Antonym Group recalled 82% of the words correctly while

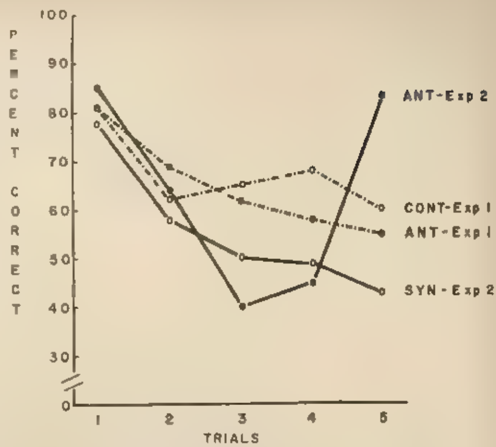


FIGURE 1. The percentage of words correctly recalled for the Antonym and Control Groups of Experiment I and the Antonym and Synonym Groups of Experiment II.

the Synonym Group recalled only 42% correctly (Figure 1). Both groups showed equivalent and significant ($p < .01$) reductions from Trial 1 recall to Trial 4 recall.

A comparison between the Antonym Group of this Experiment and the Antonym Group in Experiment I can be made since the same word triads were used on Trial 5 of both experiments. The only difference between the two groups was in the number of preceding trials that bore the antonym relationship with the word triad on Trial 5. In Experiment II four synonomous triads preceded the switch to antonyms and 82% of that critical trial was recalled correctly. In Experiment I only one such triad preceded and only 60% was recalled correctly on the critical trial. This difference was significant at the $p < .01$ level, t (98) = 2.95. Trial 4 recall performance in Experiment I was compared with the same performance in Experiment II in order to determine the difference in the extent of PI buildup in the two experiments. This difference was significant, t (98) = 2.53, $p < .05$, since *S*s in Experiment I averaged 63% (both groups combined) and *S*s in Experiment II averaged 46.5%. There were no differences between Experiments on Trial 1 performance.

DISCUSSION

The results of these two experiments, taken together, suggest that the presentation of more than one exemplar from a category may be necessary before the introduction of information from the opposite category will result in an increase in recall probability. This was shown by the fact that no increase in performance occurred when only Trials 4 and 5 contained antonyms of one another (Experiment I) but did occur when Trial 5 contained words that bore an antonym relationship with each of the preceding four trials (Experiment II). It could be that more than one instance of a relationship is needed before a subject can differentiate words that

do not participate in this particular relationship. This relationship could be established in one trial if all the words in that triad shared in the relationship, but more than one trial is needed in the present paradigm because no within triad relationships exist.

This finding poses a problem for theories that assume a word can be automatically encoded along several different dimensions immediately upon its presentation. If this were the case then the buildup of a relationship among the words should not be a necessary condition for release because differential encoding would occur as a result of automatically encoding several dimensions of each word. However, the present experiment suggests that encoding of specific features of a word may depend upon "priming" the subject for these features (Underwood, 1972). If this is true then an alternative to the automatic encoding hypothesis must be sought to explain the PI release effect.

REFERENCES

Cermak, L. S. Proactive facilitation in short-term memory. *Journal of Experimental Psychology*, 1970, 85, 305–310.
Underwood, B. J. Are we overloading memory? In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: V. H. Winston, 1972.
Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, 77, 1–15.

(Received May 12, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 742–744

SEMANTIC AND ASSOCIATIVE INTERACTIONS IN CHILDREN'S FALSE RECOGNITION[1]

PHEBE CRAMER[2] AND ANDREW P. SCHUYLER[3]

Williams College

Previous studies of memory encoding and false recognition in children have tended to focus *either* on associative relationships *or* on semantic attributes. The present study, based on the responses of third- and sixth-graders, indicated that associative and semantic factors interact to produce the false recognition effect. For Synonyms, associative strength determined the magnitude of false recognition, while for Antonyms, it was not important.

Both associative relationships and semantic features have been shown to be successful predictors of errors in children's recognition memory (e.g., Bach & Underwood, 1970; Cramer, 1972a). In general, however, those studies that have investigated the effect of associative relationships have not been concerned with the role of semantic features, while those that have investigated semantic features have not considered the possible role of associative relationships.

The influence of these two variables was investigated by Felzen and Anisfeld (1970), who compared the false-recognition errors made by third- and sixth-grade children to strongly and weakly associated synonyms and antonyms. On the basis of the finding that strong associates produced no more errors than did weak associates, they concluded that associative relationships among words played no significant role in determining such errors.

This conclusion, of course, was at variance with the results obtained in the studies cited above. Two possible reasons for the nonconfirmatory results are suggested. First, previous work with adults has indicated that there is an interaction between semantic category and associative relationship in determining recognition errors (Cramer, 1972b). If such an interaction were also a determinant of children's performance on a recognition memory task, it might mask the main effect of either variable. Second, the nonconfirmatory findings may be a function of the fact that words used to represent the various semantic and control categories were not mutually exclusive. On both lists, some of the words listed as Rhymes and nearly half of the words listed as Controls were in fact associates of the critical stimuli (based on Palermo & Jenkins, 1964, data for Grades 4–6). Considering this observed overlap, it seems likely that the other 28 filler words, interspersed throughout the continuous list presentation, might also be associatively related to the critical stimuli. Thus it is difficult to interpret the basis on which false recognition errors were being made.

The present study was designed to determine the extent to which these two factors—semantic × associative interaction and stimulus category over-

---

[1] The research reported here was part of the junior author's Psychology Honor's thesis. The authors gratefully acknowledge the cooperation of the administration and the teachers at the Williamstown Central Schools in arranging for this research to be carried out.

[2] Requests for reprints should be sent to Phebe Cramer, Department of Psychology, Williams College, Williamstown, Massachusetts 01267.

[3] Now at Cornell Medical School, New York, New York.

lap—were responsible for the failure to find strength of associative relationship as a significant factor in determining children's recognition memory. The general continuous presentation method of Felzen and Anisfeld (1970) was followed. However, great care was taken to eliminate associative overlap among the experimental, control, and filler word categories.

Method. The *S*s were 40 third-graders and 40 sixth-graders. An equal number of boys and girls was tested individually in each condition at each grade level. Following a practice list, consisting of numbers, each child was given the experimental test list. He was asked to indicate whether each item on the list was "old" or "new." The *S*s within each grade were assigned randomly within blocks of four (Weak/Strong associates $\times$ 2 test orders). Male and female *S*s were tested on alternating blocks.

In order to construct the test list, a group of 20 critical stimulus (CrS) words was selected such that half of the words elicited synonym word-association responses and half elicited antonym responses. An important criterion for the selection of these stimuli was that the associative responses should be of approximately equal strength for both third-graders (Palermo & Jenkins, 1966) and sixth-graders (Palermo & Jenkins, 1964). Of the 10 CrS–Synonym pairs, half were strongly associated (Third grade: $M = 40.0\%$; Sixth grade: $M = 47.6\%$) and half were weakly associated (Third grade: $M = 1.4\%$; Sixth grade: $M = 2.1\%$). Similarly, of the 10 CrS–Antonym pairs, half were strongly associated (Third grade: $M = 46.8\%$; Sixth grade: $M = 41.2\%$) and half were weakly associated (Third grade: $M = 1.2\%$; Sixth grade: $M = 1.2\%$). The Synonyms and Antonyms were equated approximately for Thorndike–Lorge (1944) familiarity.

From these materials, two lists were prepared. One list consisted of the 10 strongly associated CrS–Synonym and CrS–Antonym pairs; the other list consisted of the 10 weakly associated pairs. In addition, each list included 10 control words (words that had no associative relationship to any word previously presented) and 12 filler words. The words of each list were presented in two different orders, to control for possible sequence effects, making four test lists in all. Each of the four lists consisted of 59 items, and was constructed such that a word of the same class (CrS, Synonym or Antonym, Control, or Filler) was assigned the same item position on each list. Each CrS word appeared twice on the list, with 7 items intervening between the first and second presentation. Each Synonym or Antonym appeared 9 items following the second presentation of its respective CrS. A control word appeared either immediately prior to or directly after each Synonym or Antonym. The list positions not occupied by CrS, Synonyms or Antonyms, or Control words were assigned to filler words. Each list was recorded on magnetic tape at a 5-sec. rate.

Inasmuch as the study was concerned with the specific associative relationship between each CrS–Synonym or CrS–Antonym pair, it was important to make sure that no Synonym, Antonym, or Control word was associatively related to any other preceding word in the list. Thus care was taken that there was no associative relationship (associative strength = 0%) between any CrS, Synonym, Antonym, Control, or Filler word, and any subsequent experimental (Synonym, Antonym, or Control) word, other than the specific CrS–Synonym or CrS–Antonym associative relationship being investigated.

TABLE 1

MEAN NUMBER OF "OLD" RESPONSES TO SECOND PRESENTATION OF CRITICAL STIMULUS WORDS (CRS'), ASSOCIATE, AND CONTROL WORDS

	CrS's[a]	Associates[a,b]	Control[a]	Total *R*s
Third grade				
Weak list	9.0	1.6	.8	11.4
Strong list	9.1	1.5	.2	10.9
Sixth grade				
Weak list	9.1	1.2	.7	10.9
Strong list	9.3	1.6	.2	10.7

[a] Maximum value = 10.
[b] Includes Synonyms and Antonyms.

Results and discussion. The *S*s' responses were analyzed first in terms of the number of "old" responses made to CrS, Associate, and Control words. These results are presented in Table 1.

A three-way analysis of variance of these data was performed. A highly significant main effect for Word type was obtained, F (2, 152) = 1,898.45, $p < .001$. Application of the Neuman–Keuls test indicated significantly more "old" responses were given to CrS words than to Associates, and more to Associates than to Controls, all ps $< .01$. There were no other significant main effects.

A second analysis of variance, based on difference scores, considered responses to Synonym and Antonym associates only.[4] While there were no significant main effects, there was a significant Associate Strength $\times$ Semantic Category interaction, F (1, 76) = 18.40, $p < .001$. A Neuman–Keuls test indicated that significantly more responses were made to Strong Synonyms than to Weak Synonyms, to Strong Synonyms than to Strong Antonyms, and more responses were made to Strong Antonyms than to Weak Synonyms (all ps $< .01$) (see Table 2).

The results of the present experiment confirm the findings of earlier studies (e.g., Bach & Underwood, 1970; Cramer, 1972a), by demonstrating that associated words produced more false-recognition errors in children than did nonassociated words, and serve to extend those earlier findings. Contrary to the conclusions of Felzen and Anisfeld (1970), when stimulus materials were selected with adequate concern for the possibility of confounding relationships, synonyms that were strongly associated pro-

[4] Since there were twice as many opportunities to respond to Control words as there were to Synonyms or Antonyms, the difference scores for each *S* consisted of Synonym minus ½ Control and Antonym minus ½ Control.

TABLE 2

Mean Number of "Old" Responses to Synonyms and Antonyms, Corrected for Control Word Error

	Synonyms		Antonyms	
	Weak	Strong	Weak	Strong
Third grade[a]	.05	.80	.75	.55
Sixth grade[a]	.10	.70	.35	.25
Total	.15	1.50	1.10	.80

[a] Maximum value = 5.

duced more errors than did synonyms that were weakly associated.

This statement must be qualified by two further considerations. First, some other linguistic variables, not investigated in this study, may be contributing to the results. More important, from the authors' point of view, is the finding of a consistent pattern of interaction between semantic and associative factors in determining false-recognition errors. For both grades, significantly more responses were given to strong than to weak Synonyms; however, association strength was not important in determining responses to Antonyms.

Thus, the importance of association strength varies with the nature of the semantic relationship. More specifically, it may be that the semantic relationship of antonymity is particularly insensitive to associative strength. In a previous paper (Cramer, 1972b), it was pointed out that several recent studies have found associative variables to have a different, and counterpredicted effect on antonyms, as compared to other word types. Inspection of the published results of Felzen and Anisfeld (1970) reveals that the same Associative Strength × Semantic Category interaction appeared in their data.

REFERENCES

Bach, M. J., & Underwood, B. J. Developmental changes in memory attributes. *Journal of Educational Psychology*, 1970, 61, 292-296.

Cramer, P. A developmental study of errors in memory. *Developmental Psychology*, 1972, 7, 204-209. (a)

Cramer, P. Semantic × associative relationships × conditions of CRS presentation in semantic generalization. *Journal of Experimental Psychology*, 1972, 92, 246-255. (b)

Felzen, E., & Anisfeld, M. Semantic and phonetic relations in the false recognition of words by third- and sixth-grade children. *Developmental Psychology*, 1970, 3, 163-168.

Palermo, D. S., & Jenkins, J. J. *Word association norms: Grade school through college.* Minneapolis: University of Minnesota Press, 1964.

Palermo, D. S., & Jenkins, J. J. *Oral word association norms for children in grades one through four.* (Res. Bull. No. 60) University Park: Pennsylvania State University, 1966.

Thorndike, E. L., & Lorge, I. *The teacher's wordbook of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

(Received May 21, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 745–748

SHORT-TERM MEMORY LIMITATIONS ON DECODING SELF-EMBEDDED SENTENCES[1]

MAIJA S. BLAUBERGS[2]
University of Georgia

MARTIN D. S. BRAINE
New York University

Previous studies on self-embedding have not isolated the contribution of a limited short-term memory span to the difficulty of processing self-embedded sentences. With the procedure used, auditory presentation of semantically neutral sentences and the use of a memory probe technique to measure processing at five levels of self-embedding, it was found that comprehension breaks down at three degrees of embedding. A control set of sentences, semantically equivalent, but with right-branching structure, were significantly better understood at three to five degrees of complexity, but not at one and two degrees. The results are predictable, given a seven-item short-term memory span, from some plausible assumptions about the manner of processing.

An automaton with limited short-term memory cannot process sentences with unlimited self-embedding (Chomsky, 1963). Although this result must apply to humans, laboratory confirmation has proved surprisingly difficult. Miller and Isard (1964) showed that sentences of equal length and equivalent word content are memorized with increasing difficulty from zero degrees through four degrees of self-embedding. However, Blumenthal (1966) criticized this study on the grounds that naive *S*s do not process multiply-self-embedded sentences correctly, but perceive them as ungrammatical sentences with a single embedding of conjoined relative clauses modifying the initial noun. Blumenthal's *S*s were unable to correctly rewrite three-degree self-embedded sentences in a more understandable form. Since *S*s do not perceive the structure, their poor performance cannot be due to difficulty in processing the structure.

Other experimenters have found that humans can decode self-embedded sentences under suitable conditions. Schlesinger (1968) and Stolz (1967) found that they can be decoded if each predicate is pragmatically appropriate to its subject, and inappropriate to the subjects of the other clauses. Self-embedded sentences with this property were called "semantically supported." To explain his results, Schlesinger proposed a "semantic-syntactic decoding process" in which the temporary memory store merely retains the nested parts (i.e., subjects and predicates), but no order information, and then the decoder matches up the parts with the help of the semantic cues. This decoding strategy cannot be used to decode semantically neutral sentences (i.e., sentences in which any of the predicates can pragmatically be predicated of any of the subjects) because such sentences cannot be interpreted unless the order of the subjects and predicates is registered. Stolz found that semantically neutral sentences can be decoded, but only after explicit training on the self-embedded structure. However, in Stolz's study the written form of the sentence was always available as an external memory aid for the *S*s to refer to, and thus the contribution of short-term memory to the difficulty of processing self-embedded sentences was again not examined.

Since none of the studies in the literature have shown that human short-term memory limitations actually interfere with the processing of self-embedded sentences, one purpose of the experiment reported here was to demonstrate and measure the extent of the interference. While the existence of a performance deficit is certain, the size of the deficit at various levels of self-embedding depends on how humans use their short-term memory in processing sentences. Measurement of the deficit therefore provides a basis for inferences about the manner of processing.

The literature reviewed makes clear that an adequate study must present material auditorily, must use semantically neutral content (to preclude *S*s from using Schlesinger's, 1968, semantically based decoding strategy described above), and must begin by training *S*s on the nature of self-embedded sentences (otherwise errors will reflect lack of competence, not difficulty of processing). Since increasing self-embedding inevitably increases sentence length, matching right-branching sentences were used to estimate errors due to sentence length alone.

Method. Forty introductory psychology students were assigned to one of two conditions with 8 males and 12 females in each condition. In both conditions, self-embedding (SE) and right-branching (RB) training in the understanding of the sentence structure used was given, followed by tests of comprehension. The sentences varied from one to five degrees of complexity (i.e., from one to five embedded or appended relative clauses). The word content of equivalent sentences in the two conditions was the same. All sentences used in the practice and test sessions were semantically neutral (i.e., any of the subjects can form a semantically sensible sentence with any of the predicates).

[1] This paper is based on part of the first author's MA thesis at the University of California, Santa Barbara, 1970, under the sponsorship of the second author.

[2] Requests for reprints should be sent to Maija S. Blaubergs, Educational Psychology Department, University of Georgia, Athens, Georgia 30602.

Since the procedure was identical for the two conditions except for the use of SE sentences in one condition and RB sentences in the other, it will be described here only for the SE condition. The instructions read to the *S* by the *E* included an explanation of the grammatical structure of an SE sentence and an example illustrating how an SE sentence is composed of simple sentences. The explanation was accompanied by a visual display on cardboard strips of the examples used. The *S* was then given a series of practice SE sentences to read to himself. The series consisted of at least two sentences from each degree of embeddedness proceeding from one degree to five degrees. Comprehension of a sentence was tested by requiring the *S* to answer questions about the content, e.g., "Who was bickering with the Chinaman?" and "Who adored food?" were the questions for the SE sentence, "The Chinaman whom the Italian was bickering with adored food." The number of questions per sentence depended on the number of simple sentences it was composed of, i.e., there was one question per simple sentence. An exhaustive list of questions of the kind described involving all the constituent simple sentences was asked in a random order for each practice sentence. The criterion for proceeding to the next degree of complexity was that there be no more than one error in two consecutive practice sentences at each level. (This criterion was met at each level by all *S*s: SE *S*s never required more than four practice sentences to meet the criterion at any level, and RB *S*s never more than three.) The *S* was given an opportunity to correct his errors at the end of the questions for each practice sentence.

After the *S* had reached criterion at all levels of complexity, six test sentences for each level were presented by tape recorder. The order of presentation of the sentences of the differing degrees of complexity was randomized within the set of 30 sentences that each *S* heard. The *S*s were given as much time as they needed after each sentence to fill in a blank in a single simple sentence taken from the sentence they had just heard. Answers were written in an answer booklet that had separate pages for each blank. Examples of the blanks for the SE sentence given previously might be "——— adored food" and "The Italian was bickering with ———." (Only one of these sentences actually appeared in the answer booklet for this sentence.) The word to be filled in was always a noun but might be either the subject or object noun. Equal numbers of subject and object nouns (three each) were used at each level of SE, but they appeared randomly in the sequence of test sentences. The *S* was instructed not to turn the page for the next blank until after he had listened to the sentence for it, so that he did not know what information to listen for, but instead would try to comprehend the whole sentence. The *S* was allowed to hear each sentence only once and was not allowed to make any written notes about it. A practice sentence, which was reread if the *S* had any difficulty with it, was provided before the test session began. The *S* was allowed to guess when filling in the blanks for the test sentences but not told if his answers were correct or incorrect.

Results. The total number of correct responses at each level of complexity for each *S* was recorded. A correct response consisted of filling in the blank with the appropriate noun. The mean number of correct responses is plotted against the levels of complexity for both conditions (SE and RB) in Figure 1. (A chance level calculated on the assumption that the *S* would choose randomly from all nouns present in the sentence is included.)

A two-way, repeated measures, analysis of variance with *S*s nested within the two levels of the sentence-type factor (SE and RB) was used to analyze the data. The main effect of sentence type was significant, F (1, 38) = 37.4, $p < .01$, as was the main effect of sentence complexity, F (4, 152) = 134.1. The interaction between the two was also significant at the .01 level, F (4, 152) = 10.7. The significant sentence-type effect indicates that sentences with RB structure received significantly more correct responses than sentences with SE structure. The significant complexity factor indicates that level of complexity had a differential effect on the number of correct responses.

Since the interaction between the two factors was significant, the simple main effects of each factor were examined for significance at each level of the other factor. Significant differences were found for sentence type at the third, fourth, and fifth levels of complexity ($p < .01$ in all cases). Furthermore, significant differences were found for the complexity factor in each type of sentence ($p < .01$).

Scheffé tests were computed on all possible differences between pairs for the level-of-complexity factor for each sentence type. For the SE sentences, it was found that there were significant differences at the .01 level between each of the first and second levels of complexity and each of the third, fourth, and fifth levels of complexity. For the RB structure, significant differences at $p < .01$ were found between one degree and three, four, and five degrees; between two degrees and four and five degrees; and between three and five degrees.

Since the position of the test simple sentence within the longer sentence might have an effect on the number of errors, percentages of correct responses of the possible correct responses at each level of complexity for each position of a test simple sentence were calculated for the SE sentences. These percentages are as follows: 100% of responses to the first position, i.e., outermost simple sentence, and 100% of the responses to the last or innermost position, at one degree of complexity; 77.5% from the first position, 67.5% from the middle, and 95% from the innermost position at two degrees; 45% from the first, 30% from the middle, and 30% from the innermost position at three degrees; 50% from the first, 27.5% from the middle, and 40% from the innermost position at four degrees; 75% from the first, 21.3% from the middle, and 20% from the innermost posi-

tion at five degrees. The high percentages of correct responses to the outermost sentence presumably reflect combined primacy and recency effects, i.e., a tendency to remember the first subject and the last predicate. Thus, even when the entire sentence is not fully processed, the *S* may be able to correctly respond to probe items from the outermost simple sentence in the complex SE sentence. If these increased percentages of correct responses to the outermost test simple sentence at the higher levels of complexity are disregarded, it is clear that the scores at levels three, four, and five are even closer to chance level than the points in Figure 1 suggest. (A tendency for a larger percentage of correct responses to test sentences from the last or most recently presented clause in the RB sentences was also observed. Thus, the correct responses for RB sentences may also have included some responses that would not necessarily reflect processing of the complete complex RB sentence.)

The significant differences found for the complexity factor over both grammatical types support the hypothesis that comprehension decreases with degree of complexity. Furthermore, the finding of significant differences between SE and RB at the third, fourth, and fifth levels of complexity substantiates the hypothesis that the decrease in comprehension is greater in the SE condition.

Discussion. Results of the comparison between SE and RB conditions support the contention that it is complexity peculiar to SE and not merely sentence length that is causing the increased difficulty at the higher levels since there are no significant differences between SE and RB at levels one and two, but there are significant differences between SE and RB at levels three, four, and five. Length of sentence does contribute to increased errors since errors tended to increase with degree of complexity for both RB and SE sentences as shown by the downward curves in Figure 1. However, the sharp drop to near chance responses at level three for SE sentences, combined with the fact that the first significant difference between SE and RB appears at three degrees suggest that SE sentences become peculiarly difficult to comprehend at three degrees of complexity and remain so at higher degrees. Not understanding the structure has been ruled out by training the *Ss*. Length of sentence has been controlled by comparison with equivalently long RB sentences.

Thus, since the memory factor associated with length of sentence alone was constant for SE and RB sentences, the difference in errors between SE and RB sentences at each level is an unconfounded measure of increased difficulty in understanding sentences with SE structure. The reason for this difficulty is presumably the increasing demands on the short-term memory which, it is suggested here, exceed the available capacity. Thus, at three degrees of SE, in order to complete the simple sentences (i.e., recombine subjects with appropriate predicates), the decoder, hearing the predicates in reverse order to the subjects, must retain eight

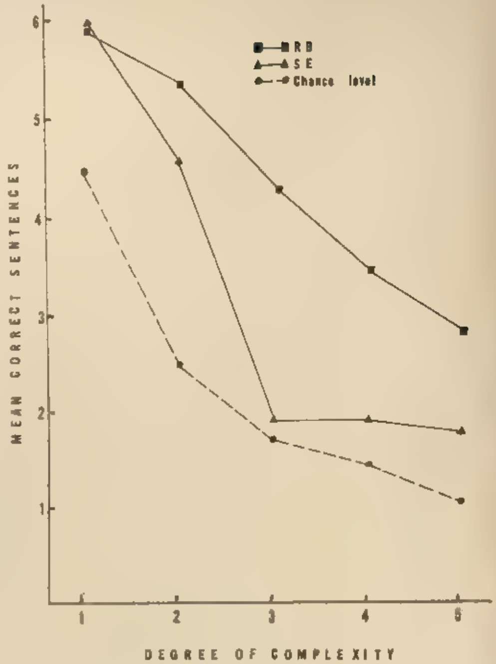


FIGURE 1. Mean number of test sentences correct as a function of complexity level. (Abbreviations: RB = right branching and SE = self-embedding.)

different items in order if recombination is done after hearing the complete sentence, or somewhere between four and eight items if recombination occurs starting with the first predicate heard. Only if the list of subjects were erased one by one as recombinations were completed before the next predicate was heard would there be only four items at a time present in the short-term memory. There is no reason to postulate such erasure. Thus if subjects and predicates are considered chunks or items of information, which seems reasonable since in the SE sentence they are presented discretely to the decoder, eight such items would have to be stored at the three-degree level. Accordingly, it is suggested that the breakdown of comprehension at the three-degree level can be accounted for by the assumption that the limitation of short-term memory to approximately seven items at a time (Miller, 1956) is first exceeded at this level of complexity. For RB sentences, the simple sentences conjoined by relative pronouns may reasonably be considered as the chunks of information to be retained. Thus, even at five degrees of complexity, a maximum of only six chunks need be retained. The results therefore support Miller's claim.

In conclusion, decoding beyond two degrees of SE for semantically neutral sentences is virtually

impossible, presumably due to structurally intrinsic short-term memory limitations.

REFERENCES

BLUMENTHAL, A. L. Observations with self-embedded sentences. *Psychonomic Science*, 1966, 6, 453 454.

CHOMSKY, N. Formal properties of grammars. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 2. New York: Wiley, 1963.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81–97.

MILLER, G. A., & ISARD, S. Free recall of self-embedded English sentences. *Information and Control*, 1964, 7, 292–303.

SCHLESINGER, I. M. *Sentence structure and the reading process*. The Hague: Mouton, 1968.

STOLZ, W. S. A study of the ability to decode grammatically novel sentences. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 867–873.

(Received May 24, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 4, 748–750

FREE RECALL OF SENTENCES AS A FUNCTION OF IMAGERY AND PREDICTABILITY[1]

P. J. HOLMES AND D. J. MURRAY[2]

Queen's University, Kingston, Ontario, Canada

Four sentences were presented for immediate free recall: the sentences were varied in imagery and in predictability. It was found that these two variables predominantly affected recall from the secondary memory component of the serial position curve, but that they exercised enough of an effect in what was ostensibly the primary memory component to suggest that even when primary memory could be used for retention *Ss* will try to process sentences in depth. In general high imagery and high predictability improved recall, with the exception that predictable abstract sentences were better recalled than predictable concrete sentences for the last serial position only.

In studies of free recall of isolated words, it has been suggested that the last three or four words presented are recalled from a short-lasting type of storage known as primary memory, whereas the early and middle items are more likely to be recalled from secondary or long-term memory (see, e.g., Glanzer & Cunitz, 1966). As yet there seems to be no comparable evidence in respect of sentence retention. It is known that two variables likely to aid sentence retention are imagery (Paivio, 1971) and predictability (Rosenberg, 1968). These variables are, however, more likely to affect secondary memory: for example, Madigan, McCabe, and Itatani (1972) showed that the imagery value of isolated words presented for free recall mainly affected the secondary memory component of free recall. It would therefore be predicted that in the free recall of sentences only the recall of the early sentences would vary with imagery and predictability. Recall of the last sentence presented, which, according to current views, would probably be recalled from primary memory, should not be affected by these two variables. The present experiment was designed to discover whether this was the case.

Method. Four types of sentence were constructed: Predictable–Concrete (PC), Predictable–Abstract (PA), Unpredictable–Concrete (UC), and Unpredictable–Abstract (UA). Twelve examples were found of each, making a total of 48 sentences. All sentences were of the form, "The (adjective) (noun) (auxiliary plus verb) (preposition) (adjective) (noun)," with the variable of imagery being manipulated via the nouns. Using the Paivio, Yuille, and Madigan (1968) ratings of noun imagery, the mean imagery values for the nouns used in the PC and UC sentences were 6.5 (range, 5.73 to 6.86) and 6.45 (range, 5.87 to 6.86), respectively. For the PA and UA sentences the mean ratings were 2.77 (range, 1.93 to 3.37) and 2.85 (range, 1.63 to 3.37), respectively. All sentences were controlled for individual word frequency using the Thorndike Lorge (1944) ratings and all sentences had approximately the same number of syllables (mean, 15.31 syllables). Sentence predictability was defined in terms of the real world likelihood of occurrence of the situation described by the sentence. Typical

[1] This work was carried out as an Honours thesis project by P. J. Holmes under the supervision of D. J. Murray. Expenses were borne by National Research Council of Canada Grant A0-126.

[2] Requests for reprints should be sent to D. J. Murray, Department of Psychology, Queen's University, Kingston, Ontario, Canada.

examples of the four types of sentence are:

PC: The enraged alligator was chasing the small boy.
UC: The angry alligator was smoking his favorite claw.
PA: The serious casualty was caused by poor [illegible].
UA: The new casualty was lacking the re[illegible]d hypothesis.

[illegible] noted that the same individual word can [illegible] both in P conditions and in U conditions: [illegible] was deliberately done in order to keep the P [illegible] U conditions as similar as possible in respect of the individual words used.

The *S*s were divided into four groups: each group received just one of the four types of sentence. Within each group, the 12 sentences were divided into three sets of 4 each, with one set of 4 sentences constituting the material for a single trial. Each *S* thus received three trials. The sentences were presented by tape recorder at a rate such that each sentence took 4 sec. to present, followed by a 4-sec. interval before the next sentence. After each trial of 4 sentences, *S* had 3 min. in which to write down the 4 sentences in any order he preferred. After he had tried to recall the sentences, he was asked to [illegible] words he might still recall.

The sentences were scored both in terms of the number of complete correct sentences, and in terms of the number of critical words retained. There were five critical words per sentence (two adjec[illegible], two nouns, one verb). Since there were few [illegible]es between the two methods of scoring, only the critical word analysis will be presented here.

The *S*s were 40 students from Introductory Psychology classes at Queen's University, with 10 *S*s in each group.

Results. Figure 1 shows the percentage of correct critical words recalled for each of the sentence types as a function of the serial position of presentation. The last sentence was nearly always recalled first, and, following the convention used typically in studies of free recall of words, the retrieval of the elements from the last sentence was considered to be from primary memory. The recall of the first three sentences was considered to be from secondary memory. It is plain from Figure 1 that the differences between the four sentence types recalled from primary memory were less than those same differences as recalled from secondary memory. An analysis of variance on the data shown in Figure 1 revealed that concrete sentences were better recalled than abstract, F (1, 36) = 66.26, predictable sentences were better recalled than unpredictable, F (1, 36) = 18.79, and that there was a strong serial position effect, F (9, 324) = 17.66. All the above values are significant at $p < .001$. The only significant interaction (at the .01 level) was Concreteness × Serial Position, F (9, 324) = 5.19: Figure 1 suggests that this interaction was due to the steeper rise of the Abstract sentences on Serial Position 4

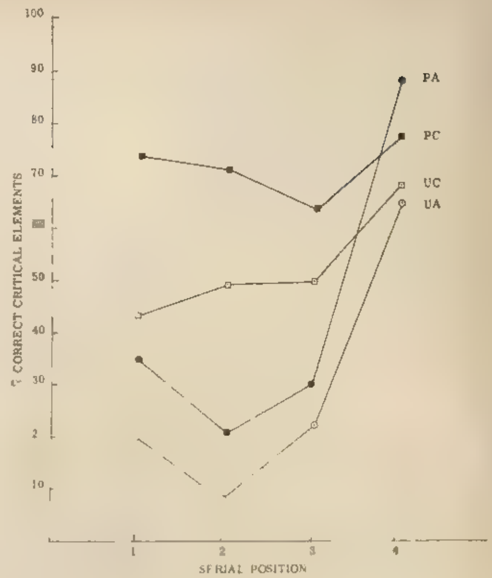


FIGURE 1. Recall as a function of imagery and predictability for the four sentences. (Abbreviations: PA = Predictable–Abstract, PC = Predictable–Concrete, UC = Unpredictable–Concrete, and UA = Unpredictable–Abstract.)

in comparison with that of the concrete sentences, and is consistent with the original hypothesis that the effects of imagery should be more marked in secondary memory than in primary memory. Nevertheless it should be noted that in primary memory the four data points shown in Figure 1 did differ significantly from each other at the .05 level as determined by Newman–Keuls tests.

A closer analysis showed that the reason the PA words were so well recalled from primary memory was that the second adjective and noun were particularly well recalled in comparison with the case for the other three groups. Since an analysis of recall by grammatical class showed that at nearly all serial positions the sentences were recalled in a "forward" manner, that is, with the first adjective and noun being better recalled than the second adjective and noun, it would seem that PA subjects paid more attention to the end of the sentences in primary memory than was normally the case. Incidentally, the fact of "forward" recall suggests that *S*s did indeed try to retain the sentences as units, rather than as collections of individual words: perhaps the first adjective–noun pair formed a "conceptual peg" on which to hook the second adjective–noun pair (Paivio, 1971). The more predictable and concrete the total sentence, the easier this would be to do.

Discussion. Imagery and predictability clearly affected recall from secondary memory, a finding consistent with the results of previous work. But in what was ostensibly retrieval from primary memory, there were two surprises: retrieval did vary

with imagery and predictability, and PA recall was best of all. Two choices are open to us: either we reject the primary–secondary memory dichotomy, at least insofar as sentence retention is concerned, or we assume that *S*s' strategies for dealing with sentences were such as to make it unlikely that the use of primary memory in sentence retention is the same as the use of primary memory for isolated words. Since Figure 1 shows that there was a clear serial position effect, we prefer to adopt the latter view. The fact that sentences were recalled in a forward manner at all serial positions suggests that, even though it was theoretically possible to store the last sentence as a collection of isolated words "echoing" for a few seconds, in fact *S*s preferred to process the last sentence in depth just as they had been doing for the earlier sentences. The advantage for the PA sentences may have resulted from an excessive concentration on the storage of the last sentence: the difficulty of the preceding three sentences may have led *S*s to compensate by storing the last PA sentence both in terms of meaning and at the same time relying on echoic memory to a greater extent than was necessary for the easier PC sentences. If this speculation is correct, it may be inferred that there is a flexibility in the use of primary memory which has perhaps been rather overlooked in the existing literature. In turn, this would suggest that some of the available evidence for primary versus secondary memory actually reflects differences in processing strategies rather than two types of storage (cf. Craik & Lockhart, 1972).

REFERENCES

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671–684.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

Madigan, S., McCabe, L., & Itatani, F. Immediate and delayed recall of words and pictures. *Canadian Journal of Psychology*, 1972, **26**, 407–414.

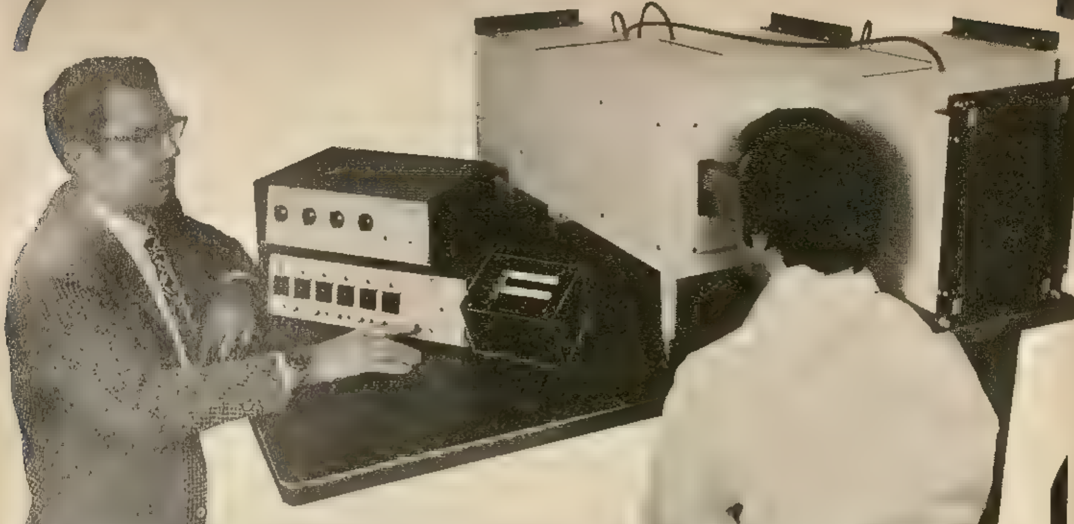
Paivio, A. *Imagery and verbal processes*. New York: Holt, Rinehart & Winston, 1971.

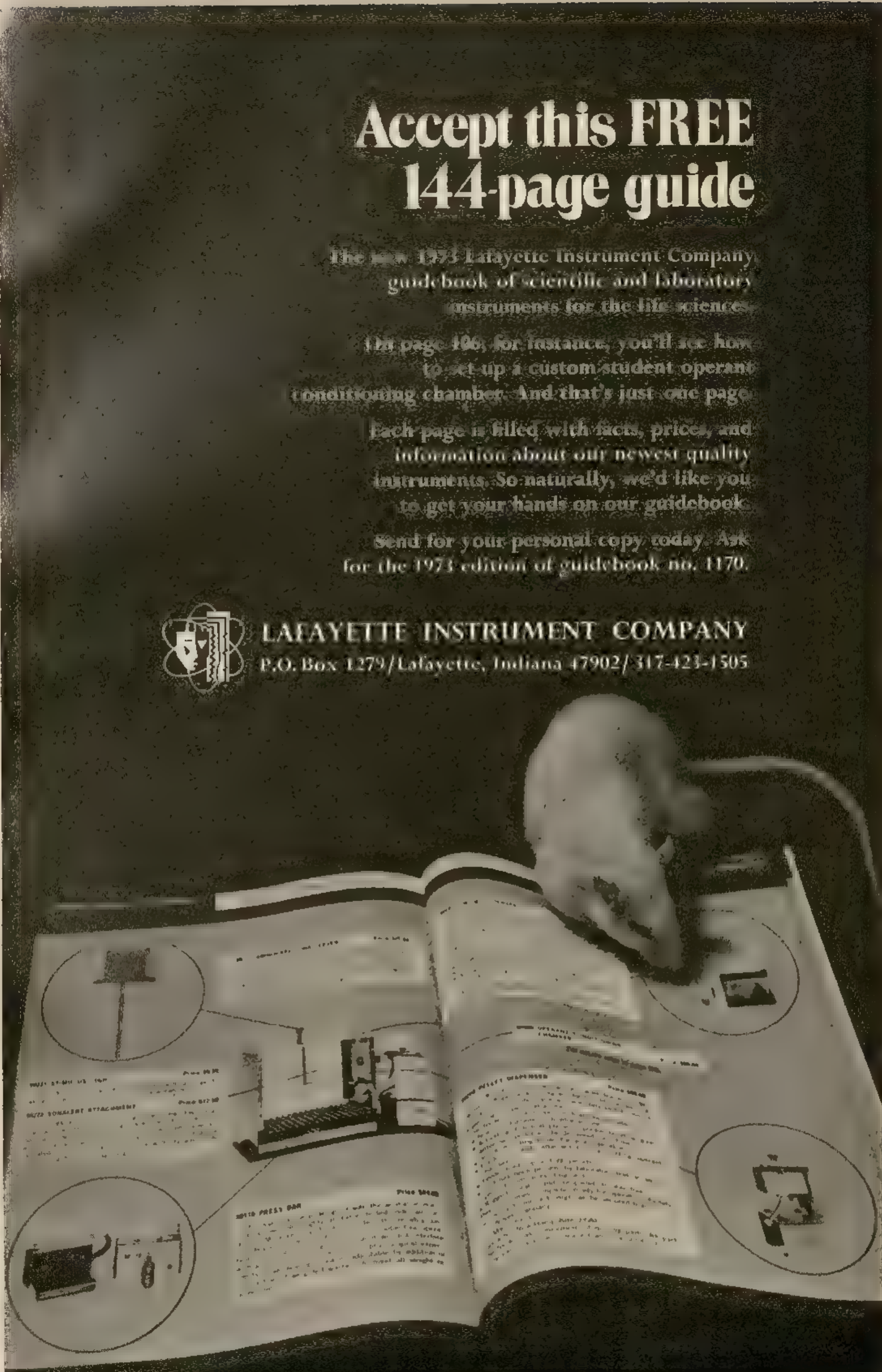
Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, Monograph Supplement*, 1968, **76**(1, Pt. 2).

Rosenberg, S. Association and phrase structure in sentence recall. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 1077–1081.

Thorndike, E. L., & Lorge, I. *The teacher's wordbook of 30,000 words*. New York: Columbia University, Teachers College, Bureau of Publications, 1944.

(Received June 25, 1973)




Accept this FREE
144-page guide
The new 1973 Lafayette Instrument Company guidebook of scientific and laboratory instruments for the life sciences.
On page 106, for instance, you'll see how to set up a custom student operant conditioning chamber. And that's just one page.
Each page is filled with facts, prices, and information about our newest quality instruments. So naturally, we'd like you to get your hands on our guidebook.
Send for your personal copy today. Ask for the 1973 edition of guidebook no. 1170.
LAFAYETTE INSTRUMENT COMPANY
P.O. Box 1279/Lafayette, Indiana 47902/317-423-1505

CONTENTS—(*Continued*)

#74

Journal of Experimental Psychology

CONTENTS

ARTICLES

(Continued on inside back cover)

American Psychological Association

Vol. 102, No. 5 May 1974

Beginning in 1975 the *Journal of Experimental Psychology* will appear as four independently edited and distributed sections:

The **Journal of Experimental Psychology: Human Learning and Memory** (Lyle E. Bourne, Jr., Editor) will publish experimental studies on fundamental acquisition, retention, and transfer processes in human behavior.

The **Journal of Experimental Psychology: Human Perception and Performance** (Michael I. Posner, Editor) will publish experimental studies designed to foster understanding of information-processing operations and their relation to experience and performance.

The **Journal of Experimental Psychology: Animal Behavior Processes** (Allan R. Wagner, Editor) will publish experimental studies concerning the basic mechanisms of perception, learning, motivation, and performance, especially as revealed in the behavior of infrahuman animals. The studies should make significant contributions to general behavior theory.

The **Journal of Experimental Psychology: General** (Gregory A. Kimble, Editor) will publish experimental articles in any of the three areas represented by the above sections, when the articles involve a longer, more integrative report leading to an advance in knowledge that is judged to be of interest to the entire community of experimental psychologists. The section will include but not be limited to articles such as those that have appeared as *Journal of Experimental Psychology* monographs and as chapters in contemporary books of "advances." Republication of a limited number of data may be permitted if necessary to make the article complete and definitive.

In some instances it may be difficult for the author to decide whether one of the above listed sections or the *Journal of Comparative and Physiological Psychology* is most appropriate for a given manuscript. The author should weigh the essential implications of the research rather than the specific techniques or subject population employed. Is the central message concerned with the issues and concepts of behavior theory (*JExP*) or with the biological substrates of behavior (*Journal of Comparative and Physiological Psychology*)? Is the study primarily addressed to theory of human behavior (*JExP: Human Learning and Memory* or *JExP: Human Perception and Performance*) or to theories equally or more relevant to infrahuman subjects (*JExP: Animal Behavior Processes*)? Is the major emphasis on the acquisition, retention, or transfer of new information, concepts, or problem-solving strategy (*JExP: Human Learning and Memory*) or on the organization or control of structures or mental operations, where the acquisition of new skills is only incidental (*JExP: Human Perception and Performance*)?

Manuscripts should be sent directly to the Editor of the relevant section at the address listed below. In cases where an Editor believes a manuscript is clearly more appropriate for an alternative journal of the American Psychological Association, he may redirect the manuscript with the approval of the author. A final action taken by one of the Editors will be final for all sections of *JExP*.

Manuscripts should be submitted in triplicate and according to the instructions contained in the *Publication Manual* of the American Psychological Association.

Authors should note that limitations on total pages make it necessary to restrict publication to those papers that in the opinion of the Editors make a notable contribution to the literature. Preference can be expected for sets of interrelated experiments that are definitive or provocative.

Addresses for the 1975 Editors-Elect are as follows: *JExP: Human Learning and Memory*, Lyle E. Bourne, Jr., Department of Psychology, University of Colorado, Boulder, Colorado 80302; *JExP: Human Perception and Performance*, Michael I. Posner, before June 1, 1974, same as Allan R. Wagner, after June 1, Department of Psychology, University of Oregon, Eugene, Oregon 97403; *JExP: Animal Behavior Processes*, Allan R. Wagner, Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510; *JExP: General*, Gregory A. Kimble, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

Subscriptions: **Subscriptions are available on a calendar year basis only** (January through December). Nonmember rates for 1974: $60 domestic, $61 foreign, $6 single issue. APA member rate: $25. Write to Subscription Section, APA.

Reprints: **Authors may order reprints of their articles from the printer when they receive galley proofs.**

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: **For information regarding microform editions write to any of the following:** Johnson Associates, Inc., 175 Fifth Avenue, New York, New York 10010; University Microfilms, Ann Arbor, Michigan 48106, or Princeton Microfilms, Princeton, New Jersey 08540.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. Undelivered copies resulting from address changes will not be replaced; second-class mail is not forwarded unless subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Published monthly in two volumes, six issues per volume, by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Washington, D.C., and additional mailing offices.

7.8.74

Journal of Experimental Psychology

Vol. 102, No. 5 — May 1974

ATTENUATION OF BLOCKING WITH SHIFTS IN REWARD: THE INVOLVEMENT OF SCHEDULE-GENERATED CONTEXTUAL CUES [1]

JAMES H. NEELY AND ALLAN R. WAGNER [2]

Yale University

Responding to an element, X, of a stimulus compound, AX, which signaled the availability of reward, was investigated in four discrete-trial bar-press experiments. "Blocking" of the acquisition of responding to X was produced in each case for some groups by separately training A alone as a signal for reward. The principle issue was whether or not blocking would be attenuated, i.e., whether or not there would be more responding to X if the reward conditions associated with AX were less favorable than those associated with A alone, as was reported by J. M. Feldman in apparent conflict with the R. A. Rescorla and A. R. Wagner model. Blocking was attenuated in Experiments I and II when the reward manipulation involved the percentage of rewarded trials and the magnitude of reward, respectively, and the A-alone training preceded AX training. Blocking was not attenuated in Experiments III and IV, which involved the same reward manipulations as Experiments I and II, respectively, but the A-alone training was intermingled with AX training. The results were interpreted as suggesting that contextual stimuli generated by the prevailing reward schedule may need to be taken into account in extension of the Rescorla Wagner model of Pavlovian conditioning to instrumental learning.

The so-called "blocking effect" occurs in Pavlovian conditioning (e.g., Kamin, 1969) when reinforced training to an elementary conditioned stimulus (CS) (e.g., A) is administered prior to reinforced training to a compound (e.g., AX) formed from that element and an added cue. The essential observation is that *S* will then be less likely to evidence conditioned responding to the added cue, X, than if *S* had received only the compound training: Pretraining to A somehow "blocks" the conditioning that would otherwise have occurred to X as a result of the compound trials. The same effect can be demonstrated when reinforced A trials are intermingled with, rather than entirely preceding, the reinforced AX trials. That is, *S* will again respond less to X alone than if training involves only the AX compound (Wagner, 1969).

To account for this effect, as well as a number of related observations, Wagner and Rescorla (e.g., 1972) have proposed

[1] This research was supported in part by National Science Foundation Grant GB-30299X to the second author and was conducted while the first author was a National Science Foundation predoctoral fellow.

[2] Requests for reprints should be addressed to Allan R. Wagner, Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510.

that the change in signal value which accrues to any cue as a result of a Pavlovian training trial depends upon the aggregate signal value ($\bar{V}$) of all the CSs present on that trial, in relationship to the asymptotic signal value (λ) that can be maintained by the consequent reinforcement or nonreinforcement which occurs. Specifically, it is assumed that the change in signal value to the ith cue in a compound of n cues may be represented as

$$\Delta V_i = \alpha_i \beta_j (\lambda_j - \bar{V}),$$

where $\bar{V} = \Sigma V_i$.

In this formulation, the values of the two rate parameters α and β are restricted to the unit interval, the former being determined by the characteristics of the specific component cue involved, the latter, like the value of λ, being determined by the characteristics of the reinforcing event. Important for present consideration is the implication for the blocking paradigm that when cue A is reinforced alone it will serve to increase the value of V_A and hence $\bar{V}$ appropriate to subsequent AX trials, and thus decrease the discrepancy, $\lambda - \bar{V}$, determining the further increment in signal value that may be enjoyed by either A or X as a result of a compound trial.

In fact, the model leads to a richer set of deductions within the general blocking paradigm if one relaxes the condition that the unconditioned stimulus (UCS) magnitude or intensity be the same on A and AX trials. For example, if the UCS (and hence λ) is less on A than AX trials, V_A cannot as nearly approach the λ appropriate to the AX trials, so that greater increments will be allowed in V_X on such trials. That is, blocking should not be as complete. Alternatively, if the UCS is *greater* on A than AX trials, blocking should either be more complete than would be the case if the UCS were the same, or V_X should become negative so that X will act as an inhibitory cue. The latter case should result when the A-alone training is sufficient to make V_A greater than the λ appropriate to the AX trials. Then $\lambda - \bar{V}$ will be negative on such compound trials, resulting in a decrement in both V_A and V_X, necessarily driving V_X negative. (See Rescorla and Wagner, 1972, and Wagner and Rescorla, 1972, for more detailed derivations.)

There are now several studies relevant to these predictions. The majority have employed conditioned emotional response (CER) conditioning in rats, and these are generally supportive of the model. Kamin (1969), for example, first paired Cue A with a 1-ma. shock and then the compound, AX, with a 4-ma. shock. This treatment produced more conditioned suppression to X alone (less blocking) than comparison treatments in which A and AX training involved equivalent shock levels, either high or low. Kamin also reported an attenuation of blocking when A was trained with a single UCS pulse, whereas subsequent AX trials involved a double pulse.

Using the same general procedures, Kamin (personal communication, June 1973) has also investigated the blocking paradigm under conditions involving a shift in UCS intensity between the component and compound occasions, but when the shift involved a reduction in intensity, i.e., when A was paired with a more intense shock than was AX. Since the obtained blocking was nearly complete in the comparison condition without a shift in UCS intensity, the study was insensitive to the prediction of the model that *greater* blocking (or inhibitory learning) should occur under the shift conditions. Important, however, was the fact that no attenuation of blocking was detected as was seen in the previously cited study (Kamin, 1969) when A was paired with a less intense shock than was the AX compound.

Rescorla, in a CER experiment reported by Wagner and Rescorla (1972), followed each AX compound by a .5-sec. shock. Intermingled A presentations were either paired with the same .5-sec. shock, in a standard blocking procedure, or were paired with a longer and presumably more aversive 5.0-sec. UCS. Subsequent testing of X, which involved transfer to a conditioned-inhibition training schedule, revealed that X had acquired less of an excitatory tendency (or more of an inhibitory tendency) in the treatment in which A alone was

ed with the stronger UCS. Wagner Rescorla interpreted this finding as ing provisional support for their model proceeded to point out that the con- ional technique (e.g., Pavlov, 1927) aking a cue, X, a conditioned inhibi- viz., reinforcing A and nonreinforcing is but the limiting case of the generally uate condition specified by their model, is, the greater reinforcement of A than

In contrast to the relatively consistent utcomes of the available CER studies is n investigation by Feldman (1971) involving a discrete-trial food-rewarded bar-press situation. In brief, it appears that in such situations a change in reinforcement between A and AX occasions may uniformly attenuate blocking, whether A or AX is the more favorably reinforced. Although the Rescorla and Wagner (1972) model is cifically addressed to Pavlovian rather n instrumental learning situations, the eneral pattern of findings concerning overhadowing, blocking, inhibitory learning, , which the model was calculated to duce, have otherwise been indiscrimina- in the two situations (e.g., Wagner, 1969; Wagner, Logan, Haberlandt, & Price, 1968). Thus, it is important to attempt to reconcile the apparent conflicts that the Feldman data occasion both with respect to the Rescorla and Wagner model and the available CER studies.

The four experiments reported here were designed to comment upon the generality of Feldman's (1971) problematic finding that blocking is attenuated when A is *more* favorably reinforced than is AX. Experiment I was an attempt to replicate the finding using the same reinforcement parameter as did Feldman, viz., different percentages of reinforced trials. Experiment II was similarly designed but employed variations in reward amount rather than percentage. Experiments III and IV used the same reward parameters as Experiments I and II, respectively, but involved an intermingled presentation of A and AX trials as opposed to successive phases of training. As will be pointed out, the pattern of findings suggests a modification in application of the Rescorla and Wagner (1972) model to account for variation in the effects of shifting reward parameters in different situations.

Experiment I

Feldman (1971) ran six groups of rats in a bar-pressing situation in which reinforcement availability was announced by either an isolated cue, A, or a compound, AX. For three of the groups the AX compound was associated with 100% reinforcement (AX-100) while for the remaining three groups it was associated with 25% reinforcement (AX-25). In either condition one of the three groups was pretrained on A with 100% reinforcement (A-100), another with 25% reinforcement (A-25), while the third, control, group was not pretrained.

In test, the responding to X alone showed the usual blocking effect when the reinforcement percentage was the same on A and AX occasions, i.e., in spite of the equivalent reinforcement of X, the A-25/AX-25 group responded less to X than the group receiving only AX-25, while the A-100/AX-100 group responded less to X than the group receiving only AX-100. The novel observation, however, was that the two pretrained groups that experienced a shift in reinforcement percentage between A and AX occasions (A-25/AX-100 and A-100/AX-25) showed more responding to X (less blocking) than did the aforementioned blocking groups that received the equivalent reinforcement percentages on AX trials.

The Rescorla and Wagner (1972) model can account for the fact that more responding to X was observed in the A-25/AX-100 group than in the A-100/AX-100 group. It does not anticipate, however, that more responding to X would be observed in the A-100/AX-25 group than in the A-25/AX-25 group. As long as the λ associated with nonreinforcement is less than the λ associated with reinforcement, it should always be the case that after an equal number of trials the signal value of a continuously reinforced cue will be greater

than the signal value of a partially reinforced cue. Thus, on reinforced AX trials, V_X should have been incremented less in the group for which A was pretrained with 100% reinforcement than in the group for which A was pretrained with 25% reinforcement. Likewise, on nonreinforced AX trials, V_X should have been decremented more in the group for which A was pretrained with 100% reinforcement than in the group for which A was pretrained with 25% reinforcement. The overall effect predicted by the Rescorla and Wagner model is that there should have been *less*, rather than more, responding to X alone in the A-100/AX-25 group than in the A-25/AX-25 group.

Experiment I was designed to evaluate the reproducibility of those aspects of the Feldman (1971) data embarrassing to the Rescorla and Wagner model.

Method

Subjects. The *S*s were 18 experimentally naive male albino rats weighing 275–375 gm. at the start of the experiment. Each *S* was individually housed with free access to water and was maintained at 75%–80% of its original weight via limited feedings given immediately after each experimental session. Two *S*s died, and 1 *S* was discarded for failure to respond during shaping, leaving 15 *S*s from which data will be reported.

Apparatus. Experimental sessions were conducted in one of six identical 24.2 × 32.4 × 18.7 cm. operant chambers housed in sound-attenuated enclosures. A ventilation fan produced a continuous masking noise of 75 db. (General Radio sound-pressure meter, 20-kHz. scale), while a .143-w. jeweled house light located directly above the food well, in the center of an end wall, provided a low level of illumination. The manipulandum was a lever located 8.3 cm. above the floor and 2.0 cm. to the left of the food well.

A 1,000-Hz. 80-db. tonal cue was provided by a Foringer stimulus generator (Model 1166-4-M1) and was delivered through a 7.5 × 14.0 cm. speaker located under the level of the grid floor directly below the food well. A diffuse visual cue was provided by a shielded 3-w. lamp located at the ceiling. When appropriate to the reinforcement contingencies, a .045-gm. Noyes food pellet was delivered to the food well from a Davis Model PD-109A feeder. All experimental events were controlled by electromechanical equipment in an adjacent room.

Experimental design. Three conditions of training equivalent to those used by Feldman (1971) were employed to evaluate the effects of reducing reward percentage between the A and AX occasions. Group AX-25 was trained only with the AX compound, signaling reward on 25% of the trials. Groups A-25/AX-25 and A-100/AX-25 received an equivalent phase of AX training but were first pretrained with Cue A alone signaling reward on either 25% or 100% of the trials as indicated. Following *S* loss, five *S*s remained in each group. For all *S*s the auditory stimulus was designated as A and the visual stimulus as X.

Shaping. Over the first five days of the experiment, *S*s were trained to bar press for food reward and gradually adjusted to a trial regime. On each of Days 1–3, *S* was exposed to four cycles of a continuous reinforcement and extinction schedule consisting of 13 min. in which either A or AX was present (depending upon which cue was first scheduled in training) and each bar press was rewarded, followed by 2 min. in which no cue was present and bar pressing was not rewarded. On Days 4–5, a trial procedure was initiated wherein the appropriate cue was briefly presented at regular intervals during 1-hr. sessions, signaling the only occasions in which reward was available. If a response occurred, the stimulus was terminated with reinforcement. In the absence of a response, the stimulus was automatically turned off, initially after 15 sec., but with a gradual shortening to 6 sec. The intertrial interval, as measured from the onset of adjacent programmed trials, was originally 15 sec. and was gradually increased to 30 sec. by the end of Day 5. Finally, a pretrial period was prescribed, just prior to each scheduled cue presentation, such that a response in this interval would cancel the trial occurrence. The pretrial duration was initially .5 sec. and was gradually increased to 5.0 sec. Thus, by the end of shaping, the conditions were in effect which prevailed throughout training—a 5-sec. pretrial period, a 6-sec. (maximum) A or AX presentation, and a 30-sec. intertrial interval.

Element training. Groups A-25/AX-25 and A-100/AX-25 received six days of training with the A cue alone prior to compound training. Each daily session was scheduled to include 96 possible trial occasions. In Group A-100/AX-25, each A presentation announced the availability of reinforcement, whereas in Group A-25/AX-25, reinforcement was available on only 25% of the trials in a pseudorandom sequence.

Group AX-25 received no element training but had the shaping sessions delayed so that all three groups entered compound training at the same time in relationship to the initiation of deprivation.

Compound training. Reward training terminated for all groups with eight daily sessions with the AX compound scheduled to announce reinforcement on 25% of the trials. As in the element training, the schedule called for 96 trial occasions within each session.

Testing. Responding to X alone as well as to A and the AX compound was evaluated in a final experimental session. Following 20 AX trials with the reward contingencies still in effect, all *S*s were administered a series of 165 test trials in which a bar press terminated the stimulus but was not

rewarded. There were 55 test trials to each of A, X, and AX in a pseudorandom sequence having the restriction that successive blocks of 15 trials include 5 to each cue. In order to insure that all test trials were received, the pretrial, time-out contingency was omitted during testing.

Results and Discussion

All *S*s rapidly adopted discriminated bar pressing during the training sequence such that on the first postshaping day the percentages of trials presented (i.e., not prevented by a pretrial response) were 88.0, 85.0, and 89.0, while the percentages of those trials presented which were responded to were 90.0, 86.0, and 92.0 for each of Groups A-25/AX-25, A-100/AX-25, and AX-25, respectively. The only remarkable deviation from such performance was on the first day in which the X cue was introduced in the pretrained groups. On this day, percentage responding to the trial cue dropped for each of the 10 shifted *S*s ($p <$ .01, sign test) in comparison with the preceding day of element training, so that the mean percentage responding was 69.0 and 63.0 for Groups A-25/AX-25 and A-100 AX-25, respectively. With the exception of this transient evidence of "external inhibition," however, response levels were so uniformly high as to preclude the detection of any potential treatment differences during training.

Table 1 presents the percentage responses to each of the test stimuli for each of the three groups. Inspection of the data from Group AX-25 suggests that the visual X cue was more salient than the auditory A cue, the former being responded to on 50.2% of its test trials, the latter on only 19.6%, after equivalent compound training. In comparison, Group A-25/AX-25 responded considerably more frequently to the pretrained A cue and less to the X cue, so as to reverse the ordering of responding to the two cues. Group A-100/AX-25, in a similar comparison to Group AX-25, evidenced less of an advantage in responding to the pretrained A cue and no apparent decrease in responding to X.

Analyses of variance indicated that the differences among the three groups in responding to X were statistically reliable, F (2, 12) = 5.54, $p < .025$. Subsequent t tests, employing the overall within-groups variance as the error estimate, revealed that Group A-25/AX-25 responded significantly less to X than did either Group AX-25, t (12) = 2.91, $p < .05$; or Group A-100/AX-25, t (12) = 2.75, $p < .05$, while the latter two groups were not reliably different, $t < 1.0$.

TABLE 1

PERCENTAGE OF BAR-PRESS RESPONSES TO EACH TEST STIMULUS: EXPERIMENT 1

Group	Cue		
	X	A	AX
AX-25	50.2	19.6	73.1
A-25/AX-25	30.5	62.5	71.6
A-100/AX-25	49.1	41.4	64.0

Note. See text for explanation of group names.

Similar analyses of responding to the A cue indicated a statistically reliable difference among the three groups, F (2, 12) = 6.81, $p < .025$, with the responding of Group A-25/AX-25 being significantly greater than that of Group AX-25, t (12) = 3.69, $p < .01$. Although the responding of Group A-100/AX-25 was greater than that of Group AX-25 and less than that of Group A-25/AX-25, these differences did not reach conventional levels of statistical reliability in either case, t (12) = 1.88, $p < .10$; and t (12) = 1.81, $p < .10$, respectively. There were no reliable differences among the groups in responding to the AX compound (F = 1.09).

The major result of Experiment 1 was to replicate the Feldman (1971) finding: There was an attenuation of the blocking of responding to X when A was more favorably reinforced during pretraining (Group A-100/AX-25) as compared to a procedure involving the equivalent reinforcement of component and compound (Group A-25/AX-25).

While this finding is apparently inconsistent with the Rescorla and Wagner (1972) model, it should also be noted that a presumptive condition was not obviously met. That is, there was no evidence that pre-

training A with the higher percentage reinforcement resulted in greater signal value to A. Indeed, the test data showed somewhat less responding to A following 100% as compared to 25% reinforcement in pretraining. As testing involved an extinction series, the relative A responding of Groups A-100/AX-25 and A-25/AX-25 may of course be taken to reflect the ubiquitous "partial-reinforcement effect." But as far as the Rescorla and Wagner model is concerned, this only raises the question of how such an effect is to be explained and leaves unchanged the fact that there was no evidence that the cue with which X was trained in compound had been pretrained to a higher signal value in Group A-100/AX-25 than in Group A-25/AX-25.

Experiment II

Reduction in reward percentage on the occasion of shift between component and compound training appears to lead to a reproducible attenuation of blocking, at least in an instrumental bar-pressing situation such as that employed by Feldman (1971) and in Experiment I. The immediate question which this raises is how such an observation may be related to the contradictory evidence of Kamin (personal communication, June 1973) and Rescorla (cited in Rescorla & Wagner, 1972), obtained in a CER situation.

Other than the major paradigmatic differences between the two training situations, the one instrumental and appetitive, the other Pavlovian and aversive, there is the obvious difference in the reinforcement parameter manipulated, percentage of reinforced trials in the present situation as contrasted with the magnitude of a consistently occurring reinforcement in the Kamin (personal communication, June 1973) and Rescorla (cited in Rescorla & Wagner, 1972) studies. To evaluate this difference, Experiment II was designed with conditions analogous to those of Experiment I but with variation in the magnitude of food reward rather than percentage reinforcement.

Method

The apparatus, *S* population, and procedure, except as noted, were identical to those of Experiment I. Twelve *Ss* were randomly assigned to each of three experimental groups designated according to the number of pellets of reward consistently received for each discriminated bar press in the separate phases of element and compound training, i.e., Group AX-1, Group A-1/AX-1, and Group A-5/AX-1.

Shaping. Preliminary shaping was carried out in eight daily 30-min. sessions. On each of Days 1–4, *S* was exposed to two cycles of a continuous reinforcement and extinction schedule as in Experiment I. Over Days 5–8, *Ss* were gradually adjusted to the trial regime as in Experiment I, except that the intertrial interval was brought to 60 sec., and Group A-5/AX-1 received five pellets reward for each response to the trial cue.

Element training. Groups A-1/AX-1 and A-5/AX-1 received 20 days of training with the A cue alone prior to compound training. Each daily session was scheduled to include 16 possible trial occasions, and on Days 6, 8, 10, 12, 14, 16, and 18 the receipt of 16 trials was insured by terminating the session for each *S* only when this number had been completed. For *Ss* in Group A-1/AX-1, each trial response terminating the cue was reinforced with one .045-gm. Noyes pellet. For *Ss* in Group A-5/AX-1, each similar response was reinforced with five pellets delivered at 150-msec. intervals.

Compound training. For the final phase of training, all three groups received 20 daily sessions of 16 scheduled trials with the AX compound and one-pellet rewards.

Testing. Responding to the compound and component cues was evaluated over 4 test sessions. On Day 1, 4 rewarded AX trials were administered, then 18 nonreinforced test trials, in such an order that 2 A, 2 X, and 2 AX trials occurred in each successive block of 6 trials. On each of Days 2–4, the 18-trial test sequence was repeated to produce a total of 24 tests to each cue.

Results and Discussion

More pretrial responding occurred than in Experiment I, but again the groups did not differ reliably in this respect nor in percentage of trial responses within any phase of training. During the common compound phase, Groups A-1/AX-1, AX-1, and A-5/AX-1, respectively, received 91%, 88%, and 92% of the scheduled trials, ($F < 1$), and responded on 88%, 93%, and 92% of the occasions ($F = 1.26$).

Table 2 presents the percentage of responses to the two component cues and the compound over the test series for each of

the three groups. In general, the pattern of data repeats that of Experiment I. The X cue was more frequently responded to than A following equivalent joint training in Group AX-1. In comparison, Group A-1/AX-1 responded considerably more often to the pretrained A and less to X. Group A-5/AX-1 evidenced no greater, indeed slightly less, advantage over Group AX-1 in responding to the pretrained A cue than did Group A-1/AX-1, with no apparent comparative decrease in responding to X.

Analysis of variance indicated that the overall difference among the three groups in responding to X was statistically reliable, $F\ (2, 33) = 3.64$, $p < .05$, while subsequent contrasts revealed that Group A-1/AX-1 responded significantly less to X than did either Group AX-1, $t\ (33) = 2.54$, $p < .05$, or Group A-5/AX-1, $t\ (33) = 2.04$, $p = .05$, with the latter two groups not reliably different ($t < 1$). Similar analysis of responding to A indicated a statistically reliable difference among the three groups, $F\ (2, 33) = 12.88$, $p < .001$, with the responding of Group AX-1 being significantly less than that of either Group A-1/AX-1, $t\ (33) = 4.65$, $p < .01$, or Group A-5/AX-1, $t\ (33) = 4.06$, $p < .01$, which themselves did not differ ($t < 1$). There were no reliable differences among the groups in responding to AX ($F < 1$).

It appears that the attenuation of blocking reported by Feldman (1971) and in Experiment I, when the pretrained A cue was more favorably rewarded than the subsequent AX compound, is not peculiar to variation in reward percentage but extends as well to variation in reward magnitude.

TABLE 2

Percentage of Bar-Press Responses to Each Test Stimulus: Experiment II

Group	Cue		
	X	A	AX
AX-1	45.8	28.5	75.0
A-1/AX-1	26.8	66.7	67.7
A-5/AX-1	42.1	61.8	74.7

Note. See text for explanation of group names.

Experiment III

The results of Experiments I and II are apparently inconsistent with the results obtained by Kamin (personal communication, June 1973) and Rescorla (cited in Wagner & Rescorla, 1972) in CER conditioning and correspondingly at variance with the Wagner and Rescorla model which anticipates the CER data. In attempting to reconcile these differences, it may be helpful to recognize that there are other major effects of manipulating reinforcement percentage and magnitude that are more typical of instrumental than Pavlovian conditioning and also not obviously accounted for by the Wagner and Rescorla model. It is possible that a manner of application of the model that would embrace these major reinforcement effects would also accommodate the blocking data of present concern.

One of the most reproducible and robust phenomena in instrumental learning situations is the partial-reinforcement effect, i.e., the fact that reinforcing only a portion of the trial responses produces more persistent responding during an extinction series than does consistent reinforcement (e.g., Capaldi & Hart, 1962; Wagner, 1961). In comparison, in studies involving Pavlovian conditioning of infrahuman *S*s, the partial reinforcement effect has frequently not been obtained (e.g., Thomas & Wagner, 1964; Vardaris & Fitzgerald, 1969; Wagner, Siegel, & Fein, 1967) or has been relatively evanescent (e.g., Wagner, Siegel, Thomas, & Ellison, 1964). Likewise, whereas it is a common observation in instrumental-learning situations that small reward magnitudes during acquisition produce more persistent extinction responding than do large reward magnitudes (e.g., Hulse, 1958; Wagner, 1961), such an effect has not been observed in comparable Pavlovian investigations (e.g., Wagner et al., 1964).

Without added assumptions, the Rescorla and Wagner (1972) model would predict extinction performance more in line with that typically seen in Pavlovian conditioning as opposed to instrumental learning. Assuming that λ has a low (e.g., zero) value

on nonreinforced trials and increases with increasing magnitudes of reward (Wagner & Rescorla, 1972), the greater the percentage or magnitude of reward, the greater should be the effective limit toward which conditioning can proceed and hence the greater the signal value from which extinction is begun.

A common theoretical device for accounting for the partial-reinforcement effect and the slower extinction following smaller rewards is to assume that the reinforcement schedule, be it in acquisition or extinction, generates stimuli that function in compound with the nominal cues. Thus, for example, Capaldi (e.g., 1966) emphasizes the stimulus "aftereffects" of reward and nonreward, and Amsel (e.g., 1962) emphasizes the cue properties of "anticipatory reward" and "anticipatory frustration." Whatever the specific identification of the schedule-generated stimuli, the argument is commonly made that shifting from an aquisition schedule to an extinction schedule produces less of a change in stimulation for partially reinforced *S*s than continuously reinforced *S*s and less of a change for small-reward *S*s than for large-reward *S*s. Thus, for example, partially reinforced *S*s are assumed to be more resistant to extinction than continuously reinforced *S*s, not because of a "stronger" response tendency, but because they have presumably been trained to respond in the presence of cues that more closely approximate those that occur during extinction.

Such reasoning, of course, is perfectly compatible with the Wagner and Rescorla (1972) model. It simply leads to an enriched conceptualization of the stimuli that must potentially be taken into account in application of the model in any situation. Thus, for example, it might be assumed that when a nominal cue, A, is rewarded on a 25% schedule during acquisition, the functional stimulus should better be represented as AS, with the latter term standing for those stimuli generated by the particular partial reinforcement schedule.[3] Rescorla and Wagner (1972) in several applications of their model have made important appeal to "situational" cues acting in compound with the nominal CS. It apparently should further be acknowledged that certain of these situational cues are a consequence of the prevailing reward schedule. To handle the fact that such extinction phenomena as the partial-reinforcement effect, which are presumed to depend upon the influence of schedule-generated stimuli, are more prevalent in instrumental as compared to Pavlovian conditioning, it need only be assumed that the relative saliences (α values) of the nominal and schedule-generated cues are likely to differ in the two situations under conventional experimental arrangements such that the schedule-generated cues typically come to carry a greater relative signal value in instrumental as compared to Pavlovian conditioning.

This formulation has important implications for cases in which reward parameters are shifted in a blocking paradigm, as in Experiments I and II, if it is assumed that there are, in fact, salient schedule-produced cues that must be taken into account. That is, it must then be assumed that the reward shift produces not only a shift in λ but a change in the stimuli present during the pretraining and compound phases. Such stimulus change should consistently work to attenuate blocking of the added cue regardless of whether the reward shift is toward a higher or lower value. It may be seen that if *S* is pretrained with AS′ rewarded and then with ASX rewarded, the degree of blocking to X will be determined by the signal value which is held by AS as a result of AS′ training, and this will be limited by differences in S and S′ which make V_S less than $V_{S'}$. When AS′ is associated with a less favorable reward schedule than ASX, both the lower λ during pretraining and the change in cues between

[3] This notation is deliberately noncommital concerning the specific source and character of the schedule-generated stimuli since such is not important here. It would be anticipated, however, that in more precise treatment it would be necessary to acknowledge that the schedule-generated stimuli are not constant over a series of trials, but change systematically with each reinforcement and nonreinforcement occasion, with time since such occasions, and likely with other variables as well (see, e.g., Capaldi, 1966).

the former and latter occasions should work to minimize the signal value of AS and allow acquisition to X, i.e., to attenuate blocking. When AS′ is associated with a more favorable reward schedule than ASX, the greater λ during pretraining and the shift in stimuli should have opposing influences upon the subsequent acquisition to X: The greater λ should work to produce a greater signal value of AS so as to minimize acquisition to X (or produce inhibitory learning), while the shift in stimuli should serve to minimize the signal value of AS and allow acquisition to X. According to this treatment, it would be expected that increasing reinforcement between pretraining and compound training would consistently attenuate blocking, whether it be in CER conditioning (e.g., Kamin, 1969) or discriminated bar-press learning (Feldman, 1971), whereas the results of decreasing reinforcement would be expected to be more variable, depending upon the relative magnitude of the opposing λ and stimulus-change effects in different experimental situations. In this light, it is significant that when decreasing reinforcement has been found to attenuate blocking it has been in instrumental reward learning (Feldman, 1971; Experiments I and II above) where there is ample other reason for assuming that schedule-produced stimuli are important controllers of behavior, rather than in CER conditioning where such an assumption is less demanded, as has been pointed out above.

This set of arguments makes analysis of the blocking paradigm relatively complicated. One must be concerned, for example, with the schedule-produced stimuli that are present during testing as well as during training. That is, it must be assumed that responding during an extinction test session is not assessed to X alone but to XS″, where S″ represents those stimuli produced by the prevailing nonreinforcement condition. Then the relevant theoretical question is not simply the degree of acquisition to be expected to X under various blocking treatments, but the signal value to be expected to XS″, where S″ can be assumed to be more similar to the stimulation produced by partial reinforcement than by continuous reinforcement, and more similar to the stimulation produced by small rewards than by large rewards (e.g., Capaldi, 1966).

As unwieldy and indeterminate as this interpretation may be, it allows at least one clear prediction. It suggests that the only factor acting to attenuate blocking when A is more favorably reinforced than AX is the change in schedule-generated stimuli between the occasions of component and compound training. Thus, if A were more favorably reinforced than AX, but in the context of the *same* schedule-generated stimuli, there should be no attenuation of blocking. Indeed, the remaining factors, such as the greater λ on A than AX occasions, should then be allowed to produce enhanced blocking or the acquisition of inhibitory tendencies to X, as otherwise predicted by the Rescorla and Wagner (1972) model.

Experiments III and IV were designed to evaluate this possibility. It is known that blocking will occur when A and AX trials are intermingled, as well as in sequential phases of training (e.g., Wagner, 1969). In such an instance, the prevailing schedule-generated stimuli which may be in compound with the nominal cues can be assumed to be equivalent on the two kinds of trials. Thus, if A were to be preferentially rewarded in comparison with AX, both kinds of trials would still occur in the context of the same schedule-generated stimuli, resulting from the intermingling of the separate rewards. In this case, there should be no attenuation of blocking as was obtained in Experiments I and II, which employed separate phases of A and AX training. Experiment III repeated the blocking comparison of Experiment I, involving different percentages of reinforcement, but with an intermingling of A and AX trials. Experiment IV similarly repeated the blocking comparison of Experiment II, involving different magnitudes of reward.

Method

The apparatus, *S* population and procedures, except as noted, were identical to those of Experi-

ment I. Six *Ss* were randomly assigned to each of three experimental groups designated according to the percentage of reinforcement scheduled in the presence of the A element and AX compound. Group AX-25 was trained only with the AX compound signaling reward on 25% of the trials. Group A-25;AX-25 and Group A-100;AX-25 received equivalent training with AX but had intermingled trials with A alone signaling reward on either 25% or 100% of the trials, respectively.

Shaping. Over the first five days of the experiment, *Ss* were gradually adjusted to a trial regime in a manner identical to Experiment I except that the cue presentations in Group A-25;AX-25 and Group A-100;AX-25 were equally distributed between A and AX, while Group AX-25 received only AX.

Training. Element and/or compound training was carried out over 14 sessions, equivalent to the total number in Experiment I, but in this case, all *Ss* received 48 scheduled AX trials daily. Group AX-25 received no further trials, while Group A-25;AX-25 and Group A-100;AX-25 received 48 additional daily trials with the A cue alone interspersed among the AX trials according to an irregular sequence.

Testing. Following 20 trials with the reward contingencies still in effect, responding to the several cues was evaluated in a final nonreinforced test series identical to that in Experiment I.

Results

As in the previous experiments, all *Ss* rapidly adopted discriminated bar pressing during the training sequence, revealing no notable differences among the three groups or between the A and AX trials within the groups receiving both cues.

The final test data of major interest are presented in Table 3. Comparing the percentage responding to X of Group A-25;AX-25 with that of Group AX-25 reveals a blocking effect similar in magnitude to that observed in the comparable percentage groups of Experiment I. The major finding of Experiment III is to be seen in

TABLE 3

Mean Percentage of Bar-Press Responses to Each Test Stimulus: Experiment III

Group	Cue		
	X	A	AX
AX-25	57.8	26.1	89.8
A-25;AX-25	37.4	76.7	79.2
A-100;AX-25	30.9	62.1	61.0

Note. See text for explanation of group names.

the relative responding to X in Group A-100;AX-25. Rather than responding similarly to Group AX-25, as did the comparable Group in Experiment I, Group A-100;AX-25 evidenced no attentuation of blocking, indeed responded somewhat less to X than did Group A-25;AX-25. Analysis of variance indicated that the difference among the three groups in responding to X was statistically reliable, F (2, 15) = 5.80, $p < .05$, with Group A-100;AX-25 as well as Group A-25;AX-25 responding significantly less than Group AX-25, t (15) = 3.25, $p < .01$; and t (15) = 2.46, $p < .05$, respectively. The difference between the two groups trained with A alone was not reliable ($t < 1$).

The pattern of responding in test to A and AX was less radically different from that of Experiment I. Again, training A alone increased the test responding to this cue in Group A-25;AX-25 and Group A-100;AX-25 as compared to Group AX-25, and somewhat less in Group A-100;AX-25, in which A alone was consistently reinforced, than in Group A-25;AX-25, in which the cue was partially reinforced. Notable also was a tendency similar to that observed in Experiment I for the group in which A alone was consistently reinforced to respond less frequently to the AX compound than did the group trained only with the compound. Separate analyses of variance performed on the responding to A and to AX each revealed significant differences among the several groups, F (2, 15) = 10.37, $p < .01$; and F (2, 15) = 5.12, $p < .05$, respectively. However, in neither case was the contrast between the two groups trained with A alone reliable. Group A-25;AX-25 and Group A-100;AX-25 both responded significantly more to A than did Group AX-25, t (15) = 4.43, $p < .01$; and t (15) = 3.15, $p < .01$, respectively, while Group A-100;AX-25 also responded significantly less to the AX compound than did Group AX-25, t (15) = 3.17, $p < .01$.

Experiment IV

The results of Experiment III, in which A and AX training trials were intermingled,

stand in contrast to those of Experiment I, in which A and AX training trials were presented in sequential phases: In Experiment I, blocking was abolished when A was more consistently reinforced than was AX; in Experiment III, blocking was not attenuated under such circumstances. Experiment IV was consequently conducted as a modification of Experiment II, involving intermingled as opposed to phased training.

Method

The apparatus, *S* population and procedures, except as noted, were identical to those of Experiment II, and when different, were patterned after those of Experiment III. Twelve *S*s were randomly assigned to each of three experimental groups designated according to the number of pellets of reward scheduled for a discriminated bar press in the presence of the A element and the AX compound during training, i.e., Group AX-1, Group A-1;AX-1 and Group A-5;AX-1.

Shaping. Over the first eight days of the experiment, *S*s were gradually adjusted to a trial regime consisting of a 5-sec. pretrial period, a 6-sec. (maximum) CS, and a 60-sec. intertrial interval, identical to Experiment II. The only change in shaping from that in Experiment II was that A and AX were equally presented from the outset in Group A-1;AX-1 and Group A-5;AX-1, whereas only the AX compound was presented in Group AX-1. As in Experiment II, the five-pellet rewards to A were delivered beyond Day 5 in Group A-5;AX-1.

Training. Acquisition was distributed over 40 sessions equivalent to the total length of element and compound training in Experiment II. All *S*s had eight AX trials scheduled daily. Group A-1;AX-1 and Group A-5;AX-1 had an additional eight A trials scheduled daily, irregularly interspersed among the AX trials. To minimize the number of trials that could be missed as a result of any pretrial responding early in acquisition, occasional sessions within the first half of training were terminated for each *S* only after the scheduled number of A or AX trials was received, in a manner similar to Experiment II.

Testing. Responding to the compound and component cues was evaluated over four nonreinforced test sessions identical to those of Experiment II. Each session included six A, six X, and six AX test trials.

Results

Acquisition performance, as in the previous experiments, was generally so high as to obscure the detection of group differences. There was, however, evidence of the higher reward magnitude in Group A-5;AX-1 as compared to Group A-1;AX-1 in a reliably more frequent responding (90.5% vs. 73.9%) to A alone over the first eight sessions, t (22) = 3.27, $p < .05$, beyond which the groups did not differ.

The percentages of bar-press responding to the several cues during testing are presented in Table 4. As expected, blocking was observed in the X responding of Group A-1;AX-1 as compared to Group AX-1. The important observation is that an equivalent blocking was seen in the X responding of Group A-5;AX-1. This finding contrasts sharply with that from the comparable reward group of Experiment II involving sequential phases of A and AX training, but is consistent with the results of Experiment III when A was also more favorably reinforced than AX in an intermingled schedule. Analysis of variance indicated a significant difference among the three treatment groups in responding to X, F (2, 33) = 6.58, $p < .005$, attributable to the lower frequency of response in Group A-1;AX-1 and Group A-5;AX-1 in comparison to Group AX-1, t (33) = 3.16, $p < .01$; and t (33) = 3.11, $p < .01$, respectively.

The lower level of responding to X in Group A-1;AX-1 and Group A-5;AX-1 in comparison to Group AX-1 was accompanied by a higher level of responding to the separately trained A cue. Analysis of variance performed on the frequencies of responding to A indicated a reliable difference among the three treatment groups, F (2, 33) = 30.98, $p < .005$, with Group AX-1 responding less than either Group A-1;AX-1, t (33) = 6.72, $p < .01$, or Group A-5;AX-1, t (33) = 6.92, $p < .01$, and no

TABLE 4
MEAN PERCENTAGE OF BAR-PRESS RESPONSES TO EACH TEST STIMULUS: EXPERIMENT IV

Group	Cue		
	X	A	AX
AX-1	59.0	23.3	77.4
A-1;AX-1	35.1	69.3	74.7
A-5;AX-1	35.4	70.8	69.5

Note. See text for explanation of group names.

reliable difference between the latter two groups. None of the differences in responding to AX approached statistical reliability.

General Discussion

In an instrumental-learning situation in which responding to one cue, A, is pretrained prior to training on a compound, AX, blocking of the acquisition of responding to X is attenuated by reducing the percentage or magnitude of reward with the introduction of the compound (Feldman, 1971; Experiments I and II above). If, however, A and AX are contemporaneously trained in the same sessions, no similar attenuation of blocking results from less favorably rewarding responding to AX vs. A (Experiments III and IV above).

Any theoretical interpretation of this pattern must somehow discount the importance of the differential rewards per se as producing the attenuation of blocking seen in Experiments I and II. The view that has been presented here is that the reward schedule, at least in some situations, contributes important contextual stimuli which work in compound with the nominal cues to control *S*'s behavior and influence the learning which occurs. Changing contextual stimuli at the time of introduction of a new cue may then be assumed to decrease the likelihood that the new cue will be treated as redundant and increase the likelihood that its signaling relationship with reward will be learned about. Such reasoning is compatible with the Rescorla and Wagner (e.g., 1972) account of the blocking phenomenon as well as alternative attentional treatments (e.g., Sutherland & Mackintosh, 1971).

The implication of the Rescorla and Wagner (1972) model remains, however, that if the contribution of the schedule-generated stimuli could be ignored, preferentially rewarding A as compared to AX should either enhance blocking or make X a conditioned inhibitor, depending upon the degree of training involved. It thus remains problematic for the theory that Experiments III and IV, which sought to create equivalent schedule-generated contextual cues during the differential rewarding of A and AX, did not provide evidence of enhanced blocking in comparison with the cases in which A and AX were equivalently rewarded.

It is not clear how seriously this apparent failure of the model should be viewed. First of all, of course, it may simply represent a detection failure. With the relatively low levels of responding to X in Groups A-25 ; and A-1;AX-1 it may be difficult to demon yet less responding in Group A-100;AX-25 Group A-5;AX-1, respectively. Indeed comparison conditions were provided to co ment upon the degree to which the respondi to X observed in these blocking groups w greater than that which would have been see to an untrained cue. More importantly, one should recognize that the designs of Experiments III and IV were such as to insure that A and AX were trained in the presence of the same schedule-generated contextual stimuli *within* any blocking group. The designs did not insure that A or AX were trained in the presence of the same schedule-generated stimuli across the different blocking groups in either experiment, or therefore that the change in schedule-generated stimuli would be equivalent in the different blocking groups when the nonreinforced test sessions were introduced. Thus, one could argue that X was indeed more effectively blocked in Group A-5;AX-1 than in Group A-1;AX-1 of Experiment IV for example, but this tendency was obscured by the differential change in schedule-generated stimuli between training and testing in the two cases. Just as clearly, however, any such arguments would be a posteriori, with the issue better decided by further investigation in which potential differences in schedule-generated stimuli might be better specified and controlled.

Additional complexities are introduced into any theoretical analysis when one assumes that schedule-generated stimuli must be taken into account along with the nominal cues. Such an assumption has been extremely useful, however, in understanding the effects of reward manipulations in instrumental-learning situations (e.g., Amsel, 1962; Capaldi, 1966; Wagner, 1961) and is likely to be as fruitful in application of the Rescorla and Wagner (1972) model to such situations, whatever the ultimate resolution of the manner in which reward shifts influence blocking.

References

Amsel, A. Frustrative nonreward in partial reinforcement and discrimination learning: Some recent history and a theoretical extension. *Psychological Review*, 1962, **69**, 306–328.

Capaldi, E. J. Partial reinforcement: A hypothesis of sequential effects. *Psychological Review*, 1966, **73**, 459–477.

Capaldi, E. J., & Hart, D. Influence of a small number of partial reinforcement training trials on

resistance to extinction. *Journal of Experimental Psychology*, 1962, **64**, 166–171.

Feldman, J. M. Added cue control as a function of reinforcement predictability. *Journal of Experimental Psychology*, 1971, **91**, 318–325.

Hulse, S. H., Jr. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *Journal of Experimental Psychology*, 1958, **56**, 48–57.

Kamin, L. J. Predictability, surprise, attention, and conditioning. In R. Church and B. A. Campbell (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts, 1969.

Pavlov, I. P. *Conditioned reflexes.* London: Oxford University Press, 1927.

Rescorla, R. A., & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black and W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research.* New York: Appleton-Century-Crofts, 1972.

Sutherland, N. S., & Mackintosh, N. J. *Mechanisms of animal discrimination learning.* New York: Academic Press, 1971.

Thomas, E., & Wagner, A. R. Partial reinforcement of the classically conditioned eyelid response in the rabbit. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 157–158.

Vardaris, R. M., & Fitzgerald, R. D. Effect of partial reinforcement on a classically conditioned eyeblink response in dogs. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 531–534

Wagner, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *Journal of Experimental Psychology*, 1961, **62**, 234–242.

Wagner, A. R. Stimulus selection and a "modified continuity theory." In G. H. Bower and J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 3. New York: Academic Press, 1969.

Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, 1968, **76**, 171–180.

Wagner, A. R., & Rescorla, R. A. Inhibition in Pavlovian conditioning: Application of a theory. In R. A. Boakes and M. S. Halliday (Eds.), *Inhibition and learning.* New York: Academic Press, 1972.

Wagner, A. R., Siegel, L. S., & Fein, G. Extinction of conditioned fear as a function of percentage of reinforcement. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 160–164.

Wagner, A. R., Siegel, S., Thomas, E., & Ellison, G. D. Reinforcement history and the extinction of a conditioned salivary response. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 354–358.

(Received September 8, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 764–767

ANAGRAM SOLVING AS A FUNCTION OF WORD IMAGERY

KATHLEEN DEWING AND PAUL HETHERINGTON[1]

University of Western Australia, Nedlands, Western Australia, Australia

Two experiments are reported which examine the role of imagery in anagram solving. In Experiment I *S*s attempted 12 anagrams constructed from nouns of high or low imagery value, with either no information, structural information, or semantic information as aids to solution. All words had a Thorndike–Lorge frequency value of 20 or more words per million, and other variables known to influence anagram solution time were controlled (word length, number of possible solutions, frequency of successive letter pairs, and letter order). High imagery words were found significantly faster than low imagery words. Both structural and semantic clues facilitated performance, but structural information had a greater effect on the solution of low imagery words and semantic information had more effect on high imagery words. It was suggested that this might reflect the use of either letter manipulation or word production strategies under the different clue conditions. In Experiment II the main effect of imagery was replicated with different word lists and additional controls. The imagery effect was again highly significant and clearly independent of word frequency and associative meaningfulness.

The image-evoking (*I*) value of stimuli is now well established as an important variable in associative learning and memory (Paivio, 1969). Imagery value is coordinated to a concrete–abstract stimulus dimension, extending from objects to pictures to concrete nouns to abstract nouns, in decreasing order of concreteness. Item concreteness or *I* value has been shown to exert an effect on paired-associate learning, recognition memory, free recall, and, under some circumstances, serial learning.

The aim of the present study was to extend the findings relating to the role of imagery in verbal learning and retrieval to the solution of anagrams. There has been some recent research interest in this area (Jablonski & Mueller, 1972), but findings have been equivocal. The effect of word imagery, if present, must occur before the word is known, as in the case of the well-established effect of word frequency on anagram solving (Dominowski, 1966; Mayzner & Tresselt, 1966).

The frequency effect has been explained in terms of the "spew hypothesis" of Underwood and Schulz (1960), which says that the order of emission of verbal units is directly related to their frequency of occurrence in the language. When the *S* looks at the letters which when rearranged will make up the solution word, it is claimed that he makes implicit word responses similar to free recall (Safren, 1962) or to free association (Mayzner, Tresselt, & Helbock, 1964) and that if the word is of high frequency, it will be more likely to be among the responses the *S* makes. If noun imagery facilitates the availability of the words as units, it would follow that anagrams of high-*I* words would be easier to solve than those of low-*I* words.

In the first experiment an attempt was made to explore the anagram solving process in more detail by providing *S*s with partial information about the solution words. The interaction between *I* value and structural and semantic clues to the solution words was examined. The structural information provided was the initial and final letter of the solution word. The semantic information was the title of the category of which the word was a member, i.e., the word's superordinate. The aim here was to determine whether the types of clues would show any differential effect on anagram solution times depending on whether the solution words were of high or low *I* value. It may be argued that if a word is of low *I* value, the likelihood

[1] Kathleen Dewing passed away since this article was written. Requests for reprints should be sent to Paul Hetherington, University of Western Australia, Nedlands, Western Australia 6009, Australia.

of its appearing among words called to mind as suggested by the letters of the anagram is lessened. The *S*s then may tend to adopt the strategy of sorting out the letters more or less mechanically until they recognize the solution word. In such cases *S*s would be expected to find a clue giving the position of 2 of the letters of the solution word of more use than semantic information.

In the second experiment an attempt was made to replicate the main effect of solution-word imagery found in Experiment I using different word lists. By assigning an estimate of A = 50 and AA = 100 to the frequency value of items, it was possible to control frequency more precisely. In this experiment the effect of verbal associative meaningfulness, *m* (Noble, 1952), was also controlled.

Experiment I

Method. A 2 × 3 repeated-measures factorial design was employed. Every *S* was given 2 anagrams to solve under each of the experimental conditions making a total of 12 problems per *S*. The independent variables were (*a*) solution-word imagery, and (*b*) type of clue given, viz., no clue, first and last letter of solution word (structural clue), or the title of the category of which it is a member (semantic clue). For example, the semantic clues for high-*I* words CABIN and STORM were DWELLING and WEATHER CONDITION, and for low-*I* words SHAME and MERCY, clues were FEELING and MORAL SENTIMENT.

The degree of association between each solution word and its semantic clue was rated on a 5-point scale by 10 independent raters without any knowledge of the purpose of the study. There was no significant difference between ratings for low-*I* words ($M = 3.8$, $SD = .72$) and high-*I* words ($M = 4.0$, $SD = .52$).

Solution words were selected from Paivio, Yuille, and Madigan's (1968) word list. Low-*I* words had an *I* value of less than 4.20 and high-*I* words a value above 6.40. All were 5-letter common nouns, and no other 5-letter common nouns could be made by rearranging the letters. All words occurred with a frequency of 20 and above per million in the language by Thorndike and Lorge's (1944) tables. Since bigram frequency (BF)—total frequency of occurrence of successive letter pairs in solution word—may be a factor in anagram solving (Dominowski & Duncan, 1964), only solution words of similar BF were included. According to tables by Underwood and Schulz (1960), fairly extreme BF values are represented by ZEBRA (1,780) and EATER (11,812). Words used in the study were all in the BF range, 3,500–8,500. The words selected were EVENT, FOLLY, MERCY, TRUTH, ARRAY, and SHAME (low-*I* words), and ANKLE, PIANO, STORM, CABIN, CHAIR, and SUGAR (high-*I* words).

Three anagrams for each solution word were constructed by 3 times randomly ordering the letters of each of the 12 words. This procedure was adopted to minimize any interfering effects arising from anagram letter order or anagram BF. The 3 sets of anagrams resulting from 3 times randomly rearranging the letters of each of the 12 words were combined factorially with the 3 clue conditions to give 9 resulting lists of anagrams, each of the treatment conditions being assigned to each anagram.

Seventy-two male and female students enrolled in first-year psychology at the University of Western Australia served as *S*s. The *S*s were randomly assigned to each of the 9 anagram lists. The nature of the task was explained, and examples given of both clue types.

The *S*s were given 3 practice trials, one with each type of clue and one with no clue. The practice trials were followed by the 12 experimental cards presented one at a time in a random order. The time taken by *S* to solve each anagram problem was recorded by *E*. To control for *E* bias, cards were placed face downward and presented to *S*s without *E* seeing the anagram until after solution times had been recorded. No provision was made to control for trial effects since it has generally been found (Mayzner & Tresselt, 1962) that *S*s do not show any great deal of improvement on such tasks until after at least several weeks' practice.

Results. Subjects were required to solve 2 anagrams under each of the 6 experimental conditions, and the *S*s' scores for each set of 2 problems were averaged. Since the distribution of response times was highly skewed, a logarithmic transformation of the form $X = \log(X_i + 1)$ was applied to the resulting data prior to within-*S* analysis of variance. The log mean solution times for word *I* values and clue conditions are shown in Table 1.

Anagrams of high-*I* words were solved faster than anagrams of low-*I* words to a highly significant degree, $F\ (1, 71) = 64.864$, $p < .001$. The variance under the 3 clue conditions was also highly significant, $F\ (2, 142) = 94.154$, $p < .001$. Interaction between *I* value and clue type was significant, $F\ (2, 142) = 5.306$, $p < .01$. This interaction includes the no-clue condition. The Semantic vs. Structural clue × *I* Value interaction was also

TABLE 1

Log Mean Solution Times for High- and Low-Imagery Words under Structural, Semantic, and No Clue Conditions

Clue condition	High imagery	Low imagery	M
No clue	1.358	1.566	1.462
Structural clue	.902	1.078	.990
Semantic clue	.704	1.182	.943
M	.988	1.275	Grand Mean = 1.132

significant, $F(1, 71) = 8.212$, $p < .01$, and the No Clue vs. Average Clue × I Value interaction was not significant, $F(1, 71) = 1.897$. Structural clues had a greater effect on the solution of low-I words and semantic information on high-I words.

Experiment II

Method. The second experiment replicated the main effect found for I in Experiment I using different words and additional controls for solution word attributes. As in Experiment I the number of letters, bigram frequency of the solution word, and number of possible solutions for each anagram were controlled. In addition, the number of redundant letters in each anagram and frequency and meaningfulness of the solution words were equated for high- and low-I lists. The words selected were POWER, CHASM, GRIEF, VENOM, CHAOS, and DRAMA (low-I words) and TRUCK, SPRAY, HOTEL, PLANK, FLASK, and DRESS (high-I words); m values used were calculated according to Paivio et al. (1968), following Noble's (1952) production method for obtaining m data. The procedure for constructing and presenting the anagrams was the same as for Experiment I, except that no clues were given. Twenty-four male and female second-year psychology students served as *S*s.

Results. Mean response time for high-I words was 32.0 sec. ($SD = 46.6$) and for low-I words 50.6 sec. ($SD = 42.4$). Anagrams constructed from high-I nouns were solved significantly faster than anagrams constructed from low-I nouns ($t = 3.5$, $p < .001$).

Discussion

The results of this study clearly demonstrate that solution-word imagery has a strong effect on anagram solving, and that this effect is independent of word frequency or associative verbal meaningfulness. Imagery must exert its influence *before* the identification of the word and in this sense displays much the same effect as frequency. The explanation that suggests itself is in similar t[illegible].

When *S*s are presented with a set of jumbled letters they at least in part recall words to match these letters (Safren, 1962). High-frequency words are more likely to be recalled by S since they are towards the top of S's hierarchy of available words. Numerous studies have found such results. The present findings suggest that the image-arousing potential of nouns also affects the availability of nouns as units. This view would predict that high-I words, because of their more ready availability, have a greater probability of being among those words suggested by the letters of the anagram. Word identification is expedited and the anagram solved more easily.

Reviews of anagram solving research indicate that both parts and wholes contribute to the solution of the problem (Johnson, 1966). Studies carried out within an S-R mediational model (e.g., Mayzner & Tresselt, 1966) have tended to restrict themselves to only one aspect of the processes involved, i.e., letter arrangement. Anagrams are treated as stimuli which evoke a variety of implicit responses in S based on successive rearrangements of the letters into new combinations, and little attention is paid to the characteristics of the word as a whole.

However, some evidence suggests that anagram difficulty in general is more [illegible] on the characteristics of the solution word than of the anagram (Dominowski, 1967). The S's solution attempt is often a word which has more, less, or other letters than does the anagram. The S jumps to his goal (word), then returns to match it against the anagram; the progression is from the word to the anagram rather than from the anagram to the word (Johnson, 1966). The availability of the solution word appears to be very important in the anagram task, and the present study provides further support for this view.

The interaction between solution-word imagery and clue type may throw some light on the strategies employed by *S*s in the solution process. Rearranging letters and producing words are different types of responses, but they are likely to occur simultaneously in anagram solving when no clues are provided, and need not be independent of each other (Dominowski, 1966). The present study showed that both semantic and struc-

tural information improved performance, but semantic clues facilitated the solution of high-*I* words more than low-*I* words and vice versa for structural clues. This may indicate that word production and letter arrangement are used differentially as strategies under those clue conditions. However, it is possible that the concrete and abstract category titles may not have been equally informative as solution clues, either because of differences in category set size or in the strength of the associative link between the word and its superordinate. An attempt was made to measure the degree of this association; the associative bond was judged by independent raters to be slightly stronger for high-*I* words, but the difference was not statistically significant. Further work is needed to clarify the issue, and the present findings should be interpreted with caution.

The results of both experiments show that imagery exerts an influence on availability of items per se and does not act purely as a mediator between stimulus and response. Experimental investigations of concreteness and the image-arousing value of stimuli have so far tended to concentrate on paired-associate learning, and have led to a rather narrow view that the effect of imagery is chiefly to cement associative relationships (e.g., Bower, 1972). The findings of the study reported here, taken in conjunction with investigations of the effects of imagery on free recall and recognition memory, suggest that imagery is much more than a mediator and may prove to be an effective variable in a wide range of cognitive tasks.

REFERENCES

Bower, G. H. Mental imagery and associative learning. In L. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1972.

Dominowski, R. L. Problem difficulty as a function of relative frequency of correct responses. *Psychonomic Science*, 1965, **3**, 417–418.

Dominowski, R. L. Anagram solving as a function of letter moves. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 107–111.

Dominowski, R. L. Anagram solving as a function of bigram rank and word frequency. *Journal of Experimental Psychology*, 1967, **75**, 299–306.

Dominowski, R. L., & Duncan, C. P. Anagram solving as a function of bigram frequency. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 321–325.

Jablonski, E. M., & Mueller, J. H. Anagram solving as a function of instructions, priming, and imagery. *Journal of Experimental Psychology*, 1972, **94**, 84–89.

Johnson, D. M. Solution of anagrams. *Psychological Bulletin*, 1966, **66**, 371–384.

Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of word transition probabilities. *Journal of Experimental Psychology*, 1962, **63**, 410–413.

Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of multiple-solution anagrams. *Journal of Experimental Psychology*, 1966, **71**, 66–73.

Mayzner, M. S., Tresselt, M. E., & Helbock, H. An exploratory study of mediational responses in anagram problem solving. *Journal of Psychology*, 1964, **57**, 263–274.

Noble, C. E. An analysis of meaning. *Psychological Review*, 1952, **59**, 421–430.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness of nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Pt. 2).

Safren, M. A. Associations, sets, and the solution of word problems. *Journal of Experimental Psychology*, 1962, **64**, 40–45.

Thorndike, E. L., & Lorge, I. *The teacher's wordbook of 30,000 words*. New York: Columbia University, Teachers College, Bureau of Publications, 1944.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.

(Received December 4, 1972)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 768–772

TRANSFER OF INFORMATION FROM SHORT TO LONG-TERM MEMORY[1]

VITO MODIGLIANI[2] AND JOHN G. SEAMON

Wesleyan University

The present study examined current hypotheses concerning information t[illegible] fer from short-term memory (STM) to long-term memory (LTM) us[illegible] Peterson STM task with word triplets presented over retention intervals o[illegible] 3, 6, 9, and 18 sec. One-third of the *S*s counted backward during the inter[illegible] one-third employed a relational imagery code, and the last third were [illegible] tracted and uninstructed in coding. The STM trials were followed by [illegible] LTM recall test of all stimuli. Half of the *S*s in each condition were forewa[illegible] of the subsequent testing. Results showed that neither warning nor dur[illegible] of information in STM had an effect on recall. Coding, however, had [illegible] nificant effect. These data were discussed in terms of the conditions whic[illegible] sufficient for storage of information in LTM.

The present study examines the conditions of information transfer from short-term memory (STM) to long-term memory (LTM) in light of current theories of memory (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). Such theories assume that information put into STM is quickly lost unless maintained by a process of rehearsal. Rehearsal may have other functions aside from maintaining information in STM. Atkinson and Shiffrin hypothesize that rehearsal "serves the purpose of increasing the strength built up in a long-term store, both by increasing the length of stay in STS (during which time a trace is built up in LTS) and by giving coding and other storage processes time to operate [p. 111]." Hence, the longer the duration of information in STM, the higher the probability of its transfer to LTM. Rehearsal thus enables information to be automatically transferred into LTM and allows time for control processes, such as coding, to be applied.

The present research is an attempt to examine the above issues within the context of the Peterson and Peterson (1959) paradigm and the extension of it by Meunier, Ritz, and Meunier (1972). Peterson and Peterson plotted the temporal course of forgetting in STM when a rehearsal-preventing distractor task was employed for 0–18 sec. Meunier et al. (1972) extended this by comparing STM performance between groups which differed by the presence or absence of a distractor task. Forgetting in STM was observed only in the group engaging in the distractor task. However, a subsequent unexpected free-recall test of all items at the end of the experimen[illegible] yielded identically poor performanc[illegible] both groups. This is consisten[illegible] view that rehearsal can ma[illegible] tion in STM, but does n[illegible] [illegible] result in transfer to LTM.

The present study [illegible] of Meunier et al. (1972) by [illegible] STM and LTM performance of [illegible] with different instructional sets in the Peterson and Peterson (1959) paradigm. The *S*s were instructed to employ a relational imagery coding strategy or a rehearsal-preventing distractor task, or were not instructed to engage in any particular activity during the retention interval of 0–18 sec. Moreover, half of the *S*s in each condition were warned of a retention test on all items at the end of the experiment.

It was hypothesized that only *S*s in the distractor group would show forgetting in the STM trials. Short-term memory performance in the coding and free groups

---

[1] This research was supported by Wesleyan University Research grants in support of scholarship awarded to both authors. Appreciation is expressed to James P. Rizza, who collected the data for this experiment.

[2] Requests for reprints should be sent to Vito Modigliani, Department of Psychology, Wesleyan University, Middletown, Connecticut 06457.

would be essentially perfect over all STM retention intervals. With regard to the more important LTM performance, it was expected that *S*s engaged in imagery coding during the STM task would show significantly better LTM performance than all other *S*s, and that *S*s warned of subsequent testing should perform somewhat better than *S*s not so warned, because of the possibility of spontaneous coding.

Method

Design. The experiment consisted of 2 phases. The first was an STM task, in which *S*s were presented with 3 word triplets to be recalled after a time interval of 0, 3, 6, 9, or 18 sec. This was followed by a recall phase, in which each *S* was asked to write down all the words he could remember.

The STM retention interval was a within-*S*s variable, while warning information and STM instructional set were both between *S*s. The experiment, therefore, can be considered a $5 \times 2 \times 3$ factorial.

Subjects. There were 10 *S*s in each of the 6 groups. They were Wesleyan University students who signed up voluntarily for the experiment and were paid $2 for their services. No attempt was made to control for sex, and *S*s were randomly assigned to experimental conditions as they showed up for the experiment.

Materials. The words used in this experiment consisted of concrete nouns with a frequency count of A or AA in the Thorndike-Lorge (1944) G count. Only nouns of 3-5 letters in length were selected to constitute the initial pool from which all successive selections were made. From this pool, 150 nouns were chosen to constitute the experimental words. They were then randomly grouped into 50 triplets with the restriction that no words in a triplet begin with the same letter and that they not be obvious associates of one another.

Each triplet was typed in capital letters on the same row, and approximately in the middle of a 12.7×20.4 cm. white index card. At the bottom of each card, a 3-digit number was also typed. The distance between the words and the number was approximately 6 cm. The 50 cards so obtained were thoroughly shuffled to yield one random order of presentation. A second order of stimuli was obtained in the same fashion. In addition to the 50 experimental cards, there were 5 warm-up cards, constructed in the same manner. The numbers on the bottom of all cards were taken from a table of random numbers.

Other materials consisted of an electric auditory clicker set at 3 clicks/sec and a screen which was put on top of a desk shielding *S* from *E*. There was an opening in the screen, slightly smaller than the cards being used, through which *E* presented all stimuli.

Procedure. Four different sequences of the STM retention intervals were used, each of which consisted of 10 blocks of 5 trials each. Each block contained a random permutation of the 5 retention intervals being used. Thus, over the 50 trials, each retention interval appeared 10 times.

Each of the 2 stimulus sequences was combined with each of the 4 STM retention interval sequences to yield 8 different sequences. Each of these was used with the first 8 *S*s in each group, with the remaining 2 *S*s given 2 of the 8 sequences selected at random.

During each trial, *S* was to read orally each of the 3 words followed by each of the digits, one at a time and in time with the auditory clicks. Reading time, determined by click frequency, took 2 sec.—1 sec. for the words and 1 sec. for the number. The *S* was explicitly asked not to pause between the words and the number, so as to minimize the opportunity for rehearsal at that point. As soon as *S* read the last digit of the number, *E* started a timer set at 0, 3, 6, 9, or 18 sec. At the end of the retention interval, *E* said "recall." The *S* was given approximately 10 sec. to respond. As soon as he responded, *E* recorded the response and then, after saying "ready," proceeded to the next presentation. The minimum time from the cue for recall to the next presentation was about 5 sec.

In the distractor condition, *S* was instructed to count backward by 3's beginning with the number he had just read. He was to count by pronouncing each digit of each number in time with the clicks until *E* gave a signal to recall the words. He was also told to avoid rehearsing the triplet and to concentrate on the counting task.

In the free condition, *S* was told that after he had read the last digit of the number there would be a period of time, varying in length from trial to trial, at the end of which *E* would give a signal for him to recall the words.

In the code condition, *S* was instructed to form an imaginary scene of the referents of the words he had just seen during the time between the presentation of the triplet and the signal for recall. The relational imagery instructions were the same as those used in a previous study (see Seamon, 1972). It was stressed that the visual scene should be one in which the single images touched each other and interacted to form meaningful, if not always realistic, units. An example of an appropriate construction, such as a car sitting in a tree with a belt around the tree, was provided to insure that *S* understood the task. No *S* reported any difficulty.

Within each of the above conditions, half of the *S*s were told that after the completion of the present task there would follow another session, in which they would be asked to recall all the words they had previously studied. For the other half of the *S*s, no mention was made of the subsequent task. Irrespective of initial instructions, all *S*s, after completing the STM phase, were given a sheet of paper and asked to write down as many of the words they had studied in the preceding trials as they could remember. They were allowed approximately 7 min. for recall, and guessing was discouraged.

Results

Responses in the STM and recall phases were scored in terms of correct triplets and correct words recalled. The latter measure was thought to be more sensitive to possible differences than the former. Triplets were scored correct regardless of word order.

In all analyses, none of the main or interaction effects associated with foreknowledge of the recall test approached significance (all ps > .10). The warning variable, therefore, was not considered further, and groups which differed only in this factor were combined.

Short-term memory retention data. Mean percentage of correct triplets and words recalled in this phase of the experiment are shown in Table 1. There was little variability associated with the code or free conditions, recall being uniformly high at all STM intervals. Recall in the distractor condition, however, whether scored for words or triplets, showed the decay function typical of this paradigm. As expected, the retention function for triplets showed a greater decrement over time than that for words, since the scoring criterion was more stringent for triplets.

Long-term memory recall data. Mean percentages of correct words and triplets recalled after the completion of the STM trials are shown in the top and bottom sections of Figure 1, respectively. Since a floor effect for the distractor and free conditions was evident when recall was scored for triplets, only recall for words was analyzed statistically. An overall analysis of variance showed that STM retention interval had a significant effect on the subsequent LTM recall of words, F (4, 216) = 14.58, $p < .01$, with Neuman–Keuls comparisons showing recall to be better at 3 sec. than at 0 sec. ($p < .01$) and at 18 sec. than at 3 sec. ($p < .05$). A significant STM Interval × Group interaction was also found, F (8, 216) = 7.27, $p < .01$, indicating that relative recall performance between groups was not constant over all STM retention intervals. The code condition was poorest at the shorter intervals, but best at the longer intervals. The main effect of group was not significant.[3]

Analyses of simple main effects followed by Neuman–Keuls comparisons for each experimental group found no significant comparisons between intervals for the free condition, significantly better recall at 3 sec. than at 0 sec. for the distractor condition ($p < .05$), and significantly better

TABLE 1

Mean Percentage of Short-Term Memory Recall for Triplets and Words

Experimental condition	Retention interval (in sec.)				
	0	3	6	9	18
Code					
Triplets	98.5	98.5	98.0	98.5	99.0
Words	99.5	99.3	99.0	99.3	99.7
Free					
Triplets	99.5	98.5	98.5	97.5	96.0
Words	99.8	99.7	99.5	99.2	98.8
Distractor					
Triplets	95.5	52.5	35.5	24.5	17.5
Words	98.8	74.2	64.5	51.5	37.5

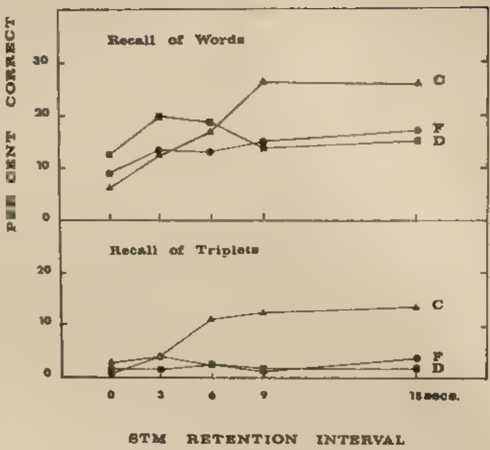


Figure. 1. Mean percentage of correct recall of individual words and triplets for the Code (C), Free (F), and Distractor (D) conditions as a function of the short-term memory retention interval.

[3] Essentially the same findings were obtained in a recognition test given to all *S*s after the recall test. Recognition performance was similar and did not increase with STM interval length for the free and distractor conditions, while the performance for the code condition was initially poorer than that of the other conditions, but improved through the 9-sec. interval to surpass them. These data do not, however, provide independent information.

recall at 3 sec. than at 0 sec. ($p < .05$) and at 9 sec. than at 6 sec. ($p < .01$) for the code condition. Comparisons of groups as a function of STM interval found significant differences at 3, 9, and 18 sec., with the distractor condition better than the code condition at 3 sec. ($p < .05$), and the reverse true at 9 and 18 sec. (both ps $< .01$). Other differences were not significant, although the distractor–code comparison approached significance at 0 sec. ($p < .10$).

The mean number of false recalls, i.e., words not from either the warm-up or experimental trials, was 2.0 across all groups, as compared to an overall mean number of experimental words recalled of 23.4.

Serial position data. The STM phase of the experiment lasted approximately 15 min. In the second phase of the experiment, therefore, some of the to-be-recalled words had been studied as long ago as 15 min. and some as recently as 2–4 min. before recall. It was desirable to examine, albeit informally, whether serial position in the STM list had an effect on recall. For this purpose, the STM list, consisting of 10 blocks of 5 triplets each, was divided into 3 groups, viz., Blocks 1–3, 4–7, and 8–10, covering the initial, middle, and end portions of the list, respectively. Two overall trends were noted. First, regardless of list position, words studied at 0 and 3 sec. were always recalled more poorly in the code condition than in the distractor condition, with the free condition intermediate. The reverse was true for words studied at 9 and 18 sec. This indicates that the recall pattern shown in Figure 1 was fairly constant over the range of possible list positions. Second, a strong recency effect was present, with much better recall from Blocks 8–10 than from Blocks 1–3 or 4–7.

Discussion

The major focus of this study was on the mechanisms of information transfer from STM to LTM. Such mechanisms will be discussed with reference to the following 4 propositions, which summarize the major tenets of current theories of memory on this subject. (*a*) Amount of information transfer is an increasing function of the time an item remains in STM (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). (*b*) Rehearsal strengthens LTM traces (Bjork, 1970). To the extent that rehearsal keeps an item in STM, this proposition is a corollary of the first. However, since rehearsal per se is currently receiving a great deal of attention, we prefer to keep these 2 propositions separate. (*c*) Rehearsal provides more time for coding processes to operate, and (*d*) memory traces in LTM can be strengthened by such coding processes.

Consistent with earlier data (Jacoby & Bartz, 1972; Meunier et al., 1972), the present findings offer no support for the proposition that prolonged stay (up to 18 sec.) in STM increases an item's probability of transfer to LTM. This is shown both by an analysis of the free condition itself, and by a comparison with the distractor condition. First, the finding of essentially perfect retention in the free condition over all intervals in the STM trials indicates that triplets were held in STM for varying intervals up to 18 sec. However, no increase in LTM retention was observed as a function of STM interval length, a result clearly at odds with the proposition being discussed. Second, while the distractor and free conditions were qualitatively and quantitatively different in STM recall, they were alike in LTM retention. The STM decay function found in the distractor condition means that items were lost from STM very rapidly. In spite of this, they were remembered as well as items which, in the free condition, had been kept in STM for longer periods of time, again at odds with the proposition in question.

With respect to the proposition that rehearsal, directly or indirectly, increases an item's strength in LTM, the situation is more ambiguous. The present data rule out the possibility that rehearsal, by virtue of simply keeping items in STM, automatically transfers them to LTM. Further, if it is assumed that *S*s engaged in rehearsal to maintain the items in STM over the retention intervals, the LTM performance of the free and distractor conditions would suggest that either rehearsal does not strengthen LTM traces, or the amount of rehearsal was the same for both conditions and over all intervals. This follows from the fact that there was more opportunity for rehearsal in the longer STM intervals than the shorter intervals and in the free condition than in the distractor condition.

Does rehearsal provide time for coding processes to operate? A comparison of the free and code conditions would suggest that spontaneous coding strategies, if employed, were not very effective in engendering LTM retention, or that *S*s simply did not engage in coding processes when given the opportunity in this task. It is still possible, of course, that under different conditions, effects of spontaneous coding would be evidenced.

We shall now consider the fourth of the propositions under scrutiny, the one concerning coding processes. As Figure 1 shows, LTM performance of the code group was poorest for words shown at 0-sec. STM intervals and increased up to asymptotic level as the STM interval increased. At 9- and 18-sec. STM intervals, this group showed better LTM performance than either the free or distractor groups. The monotonic increase in LTM retention as a function of STM interval in the code group suggests that it takes time to apply the imagery code to words, a time that can be estimated to be within 3–9 sec., judging from the recall function. This parameter may actually be expected to vary with the number and concreteness of the words employed. Moreover, it can also be expected to vary with practice. This can be inferred from the fact that the highest relative increase, as a function of trial blocks in the STM list, occurred for items shown at 0-sec. STM intervals, while the smallest relative increase occurred for items shown at 9 and 18 sec. The indication is, therefore, that *S*s became better at imagery coding with practice, and this increased efficiency was most beneficial at very short STM intervals.

The finding that the code condition, given sufficient time, produced superior LTM retention as compared to the free and distractor conditions is consistent with previous findings (Bower, 1972) indicating the superiority of the mnemonic device over other methods and with the view that organization enhances LTM retention. Thus, imagery coding is cl... sufficient condition for storage of infor... in LTM and, as judged by the lack of a w... effect, intent to remember may be superfl...

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation*. Vol. 2. *Advances in research and theory*. New York: Academic Press, 1968.

Bjork, R. A. Repetition and rehearsal mechanism in models for short-term memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Bower, G. H. Mental imagery and associative learning. In L. W. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1972.

Jacoby, L. L., & Bartz, W. H. Rehearsal and transfer to LTM. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 561–565.

Meunier, G. F., Ritz, D., & Meunier, J. Rehearsal of individual items in short-term memory. *Journal of Experimental Psychology*, 1972, 95, 465–467.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, 58, 193–198.

Seamon, J. G. Imagery codes and human information retrieval. *Journal of Experimental Psychology*, 1972, 96, 468–470.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, 72, 89–104.

Thorndike, E. L., & Lorge, I. *The teacher's wordbook of 30,000 words*. New York: Columbia University, Teachers College, Bureau of Publications, 1944.

(Received July 2, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 773–777

GOOD CONTINUATION REVISITED [1]

LUBOMIR S. PRYTULAK [2]

University of Western Ontario, London, Ontario, Canada

Twenty female *S*s inspected patterns of three lines of dots radiating from a common dot, in which one line (the shaft) was longer than the other two lines (the arms). Which arm the shaft appeared to group with, and which of two arms in different patterns the shaft grouped with more strongly, were predicted from three predictors: (*a*) rectilinearity—grouping is stronger with the arm closer to forming a 180° angle with the shaft, the function between strength of grouping and angle being nonlinear; (*b*) orientation—grouping is stronger with an arm on the vertical-horizontal rather than an oblique arm; and (*c*) enclosure—grouping is stronger when the arm closer to rectilinearity is ipsilateral with the other arm than when it is contralateral. Computer-generated predictions were correct 16/16 times when predicting grouping within individual patterns and were correct 103/109 times when predicting inequalities of grouping between patterns.

The principle of good continuation was introduced into the psychological literature in 1923 by Max Wertheimer (1958). In several demonstration displays (Figures 1–5, pp. 126–127), he showed that when three straight lines of dots radiate from a common point, the two lines forming a 180° angle appear to group more with each other than either appears to group with the third line. This is the fundamental good continuation effect which the present paper attempts to elaborate. More particularly, using three-line patterns similar to Wertheimer's, three questions are explored. First, what are the grouping effects at various angle juxtapositions, including ones that do not contain 180°? Second, does the orientation of the lines with respect to the vertical-horizontal affect grouping? Third, does ipsilateral vs. contralateral juxtaposition of two lines at constant angles to a third line affect grouping?

Method

Stimuli. Each stimulus was made up of 2-mm.-diam. circular black dots on a 21.6 × 27.9 cm. white card. The dots were arranged in three lines radiating from a single shared dot, as in Figure 1. Within each line, dots were placed at 5-mm. intervals. A line could appear at any of eight orientations—four on the vertical and horizontal and four 45° off the vertical or horizontal. One of the three lines (the *shaft*) was longer than the others—it contained four dots, counting the shared dot. The remaining two lines (the *arms*) contained three dots, again counting the shared dot. As the shaft appeared at eight orientations, and for each of these the two arms assumed $\binom{7}{2} = 21$ pairs of orientations, the number of stimuli was 168. When one of the arms came closer to forming a 180° angle with the shaft, it was referred to as Arm a and the other arm as Arm b; when the arms were symmetrical about the shaft, however, this labeling could not be used. Figure 1 shows a stimulus in which Arm a is at 90° to the shaft and Arm b is at 45° to the shaft—the "angle of an arm" always refers to the smaller angle between that arm and the shaft.

Subjects. Twenty female students at the University of Western Ontario were paid to participate in a series of experiments on grouping and visual illusions, of which the present experiment was one. At no time, however, did *E* discuss with *S* the nature of the experiments, or apprise her of data from earlier experiments or of hypotheses relevant to the current experiment.

Procedure. The *S* placed a stimulus card on a stand tilted back 15° from the vertical (where its illuminance was .088 cd/cm²), rested her chin in a chin rest (which gave a viewing distance of 45 cm.), pushed a button which turned a quietly-sounding buzzer off for 5 sec., and fixated the shared dot

FIGURE 1. Example of a stimulus display.

[1] This research was supported by Grant APA-390 from the National Research Council of Canada.

[2] Requests for reprints should be sent to Lubomir S. Prytulak, Department of Psychology, University of Western Ontario, London, Ontario N6A 3K7, Canada.

during the 5 sec. of silence. During fixation, she decided which of the two arms appeared to group with, or belong to, the shaft. She indicated her experience of grouping by writing an integer from −3 to +3 excluding 0. Positive values indicated that the shaft appeared to group with Arm a and negative values, with Arm b (a small replica of the stimulus drawn in the upper-left corner of the stimulus card labeled Arms a and b). The integers 1, 2, and 3 indicated that grouping was weak, moderate, or strong. After recording her rating, *S* found the stimulus number on the back of the card and wrote it beside her rating. Before running each *S*, the deck of cards was rearranged by shuffling as well as by dealing the cards into separate piles in haphazard order, then picking the piles up in haphazard order.

Data simplification. In order to reduce the task of data presentation and data interpretation to manageable proportions, 144 of the 168 stimuli were sorted into 18 *patterns* of 8 stimuli each. (The remaining 24 stimuli were ones in which the arms were symmetrical about the shaft and will be discussed below.) The 8 stimuli that were considered to be instances of a single pattern were stimuli in which the two arms bore the same relation to the shaft and in which the shaft was either always on the vertical-horizontal or always oblique. Below the abscissa in Figure 2, for example, Pattern 1 stands for 8 stimuli, 4 of which can be generated by rotating Pattern 1 in 90° steps, and the other 4, by rotating a mirror image of Pattern 1 in 90° steps. Similarly, Pattern 10 stands for 8 stimuli whose arms bear the same relation to the shaft as in Pattern 1, but now with the shaft oblique. The data in Figure 2 were produced by averaging together each *S*'s ratings to the 8 stimuli within 1 pattern, then calculating the means and standard errors of the means across *S*s.

What is predicted. Two sorts of predictions are made. The first is which arm within a pattern will group more strongly with the shaft. These predictions are called *within-pattern* predictions and consist merely of specifying the sign of the mean ratings.

The second sort of prediction specifies inequalities between means. These predictions are called *between-pattern* predictions. Between-pattern predictions, however, were attempted only for pairs of means originating from *comparable* patterns. Comparable patterns are pairs of patterns in which one of the arms in the first pattern is at the same angle to the shaft as one of the arms in the second pattern. These two arms are called *common* arms, and the remaining two arms are called *target* arms. Thus, 1b–2b are common arms and 1a–2a are target arms; 1a–15b are common arms and 1b–15a are target arms; and so on. In between-pattern comparisons in which the angles of the arms in one pattern are the same as the angles of the arms in another pattern (1–5, 2–4, and others), the a arms will be considered to be the target arms.

The between-pattern predictions generated below will be for the target arms of comparable displays. Thus, no prediction will be ventured for Patterns 1–9 or 1–18 because they contain no common arms and, therefore, no target arms for which predictions can be made. Of the $\binom{18}{2} = 153$ combinations of patterns, 129 proved to be comparable.

Three predictors. Three predictors were used: *rectilinearity*, *orientation*, and *enclosure*. These are described below.

The first predictor, rectilinearity, states that a shaft groups more strongly with the arm with which it comes closer to forming a 180° angle. The increase in the grouping of an arm to the shaft, however, is considered to be a nonlinear function of angle. If we start off by imagining a linear function, we can say that at every 45° increase in angle, the increase in strength of grouping is "moderate," in comparison with which any larger increase would be "large." If we now imagine that the point in the function above 135° is pulled down so as to be on a level with the point above 90°, the 90°–135° increment will be zero, and the 135°–180° increment will be large. By inference, the 45°–135° increment will be moderate, and the 45°–180° and 90°–180° increments will each be large. In short, all comparisons in which one arm is at 180° imply a strong rectilinearity effect, all comparisons in which the two arms are at 90° and 135° imply no rectilinearity effect, and all other comparisons imply a moderate rectilinearity effect.

For an example of a within-pattern prediction, we see that in Pattern 1, Arm 1a is at 90° and Arm 1b is at 45°—grouping, therefore, is predicted to be with 1a so that the sign of the mean rating should be positive. In between-pattern prediction, we might compare the two target arms 1a and 3a—3a is at 180° and is, therefore, predicted to receive a higher rating than 1a which is at 90°. (Whenever a target arm is a b arm, the sign of the mean rating for that pattern must be reversed. In predicting 1b < 15a, for example, the data are considered to show $-1.27 < -.14$.) Rectilinearity, however, is unable to generate a prediction in every instance. Thus, it generates no within-pattern prediction for Patterns 6, 8, 15, and 17 (because within each pattern the two arms are at 90° and 135°) and generates no between-pattern prediction for Patterns 1–5 or 1–10, for example (because the target arms are at equal angles to the shaft), or in the case of 1–2 or 9–16, for example (because the target arms are at 90° and 135°).

The second predictor, orientation, states that a shaft groups more strongly with an arm on the vertical or horizontal than with an oblique arm. Within Pattern 1, for example, 1a is horizontal and 1b is oblique—thus, orientation joins rectilinearity in predicting 1a > 1b. In the case of Pattern 10, however, orientation works against rectilinearity—whereas rectilinearity predicts 10a > 10b, orientation predicts that 10a < 10b. (The resolution of contradictory predictions is discussed in the next section.) In cases where both arms are oblique (like Pattern 2) or both vertical–horizontal (like Pattern 11), orientation makes no prediction.

In between-pattern comparisons, the orientation

prediction is more complex to derive. First, each target arm is assigned a value of +1, 0, or −1 depending on whether the within-pattern prediction is that the target arm is helped, unaffected, or hurt by orientation. Second, the values assigned the target arms within each pattern are compared—a shaft is predicted to group more strongly with the target arm which has the higher value, otherwise no prediction is made. For example, 1a gets +1 and 10a gets −1—therefore, orientation predicts 1a > 10a; 1a(+1) − 2a(0)—therefore, orientation predicts 1a > 2a; and 2a(0) − 11a(0)—therefore, orientation makes no prediction.

The third predictor, enclosure, is used only in between-pattern comparisons and only when the target arms in comparable displays are at the same angle to the shaft. Enclosure specifies that the shaft groups more strongly with the target arm when the target arm is ipsilateral with, and thus encloses, the common arm. Illustrative predictions are 1a > 5a and 2a > 4a.

Contradictory predictions. When only one predictor was applicable, or when two predictors were applicable and both predicted the same thing, the prediction was said to be unequivocal. (In no comparison were all three predictors applicable.) In the case of equivocal predictions, however (when two predictors were applicable and made opposite predictions), a simple tie-breaking rule was invoked —when rectilinearity contradicted orientation, rectilinearity was expected to dominate when the rectilinearity effect was large (one target arm was at 180°) and orientation was expected to dominate when the rectilinearity effect was zero (the target arms were at 90° and 135°)—although the latter case could have been classed as unequivocal, it was classed as equivocal so as to distinguish it from cases in which rectilinearity differences did not exist and rectilinearity was totally inapplicable. In all other cases of contradiction, no prediction was made—the antagonistic predictors were considered to be equivalent in strength so that no overall prediction could be ventured with safety.

Stimuli with symmetrical arms. Data from 24 stimuli will not be discussed—these are stimuli in which the arms were symmetrical about the shaft so as to form a Y, a T, or an arrowhead. They will not be discussed for four reasons: (*a*) no predictor was able to generate a prediction on the sign of the mean rating within a given stimulus; (*b*) in between-stimulus comparisons, there was no basis for deciding which arms to call common arms and which to call target arms, and thus there was no way of knowing which arm a prediction should be made for; (*c*) four, rather than eight, stimuli existed per pattern; and (*d*) no systematic tendency in the ratings given these stimuli was apparent.

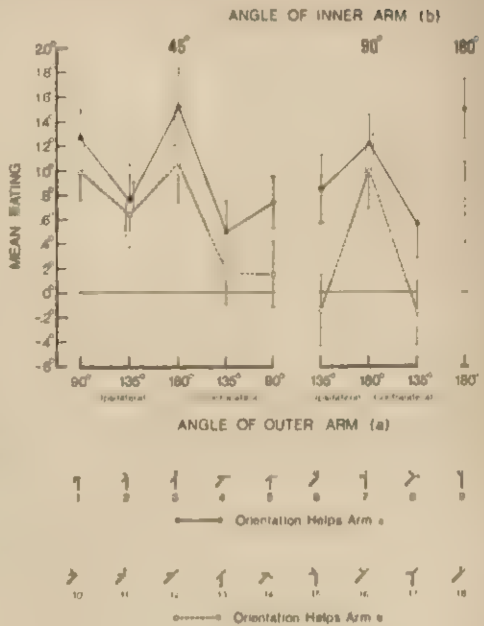


FIGURE 2. Means with ±1 standard error of each mean for 18 stimulus patterns. (An instance of one of the 8 stimuli in each pattern is shown below the abscissa in schematized form, with Arm a dashed. Positive ratings indicate that the shaft grouped with Arm a; negative ratings, with Arm b.)

RESULTS

Figure 2 is organized as follows. The two functions on the left present data from patterns in which Arm b is at 45°, with Arm a rotating from 90° ipsilateral with Arm b, to 90° contralateral to Arm b. The two middle functions present data from patterns in which Arm b is at 90°, with Arm a rotating from 135° ipsilateral to 135° contralateral. The two right-hand data points are from patterns in which Arm b is at 135° and Arm a at 180°. Finally, with the exceptions noted below, the solid functions present data from patterns in which orientation promotes grouping toward Arm a, and the dashed functions present data from patterns in which orientation promotes grouping toward Arm b.

Within-pattern predictions. Rectilinearity and orientation both predicted grouping toward Arm a in Patterns 1, 3, 5, and 9. Rectilinearity alone predicted grouping toward Arm a in Patterns 2, 4, 7, 11, 13, and 16. Orientation alone predicted grouping toward Arm a in Patterns 6 and 8 and toward Arm b in Patterns 15 and 17 (in

these four patterns, the rectilinearity difference was the impotent 90°–135°). Orientation favoring Arm b opposed rectilinearity favoring Arm a in Patterns 12 and 18, but rectilinearity was predicted to dominate because of the presence of a 180° arm. Finally, no prediction was made for Patterns 10 and 14 where orientation favoring Arm b was opposed by a moderate rectilinearity effect (arms at 45° and 90°) favoring Arm a.

Overall, then, the algorithm made predictions on the sign of the mean rating given 16 of the 18 patterns, and was correct 16/16 times, $p < .001$ in the binomial distribution $n = 16$, $p = .5$. (The binomial distribution was used to calculate all probabilities reported below.)

Between-pattern predictions. As noted above, the algorithm for generating between-pattern predictions showed that 129 of the 153 pattern combinations were comparable. In 20 of the 129, however, the algorithm generated no prediction for the following reasons. In 3 comparisons (2–11, 4–13, and 7–16), no predictor applied. As the orientation predictor was among those that failed, it was unable to determine which member in each of the three pairs should go in the solid "Orientation Helps Arm a" functions and which in the dashed "Orientation Helps Arm b" functions. Accordingly, Patterns 2, 4, and 7 were placed in the solid functions in order to keep the solid and dashed functions from crossing. In 17 more comparisons, two predictors made contradictory predictions which the algorithm was not programmed to resolve—in 2 comparisons (8–15 and 5–10), orientation opposed enclosure and in 15 comparisons, rectilinearity opposed orientation with target arms at 45° and 90° (2–6, 8; 4–6, 8; 6–11, 13; 7–12; 8–11, 13; and 12–16) or at 45° and 135° (9–12; 10–15, 17; and 14–15, 17).

Of the 109 comparisons that remain, predictions were generated and were correct 103/109 times ($p < .001$). These may be broken down into 75/79 correct ($p < .001$) when predictions were unequivocal and 28/30 correct ($p < .001$) when predictions were equivocal.

A further breakdown of the 79 unequivocal predictions shows: (*a*) 41/45 correct ($p < .001$) when rectilinearity and orientation made compatible predictions (prediction confirmed: 1–6, 7, 8, 16; 2–3, 9, 14, 15, 17; 3–4, 7, 10, 11, 13, 14, 16, 18; 4–9, 14, 15, 17; 5–6, 7, 8, 16; 6–9; 7–15, 17, 18; 8–9; 9–11, 13; 11–14, 15, 17; 13–14, 15, 17; 15–16; and 16–17, 18; prediction disconfirmed: 2–10; 4–10; 10–11, 13)—the 4 incorrect predictions, it will be noted, all involve Pattern 10, whose mean was expected to be lower, even lower than the mean of Pattern 13; (*b*) 18/18 correct ($p < .001$) in comparisons in which rectilinearity was the only applicable predictor (1–3, 15, 17; 3–5, 9; 5–15, 17; 6–10, 14, 18; 8–10, 14, 18; 9–15, 17; 10–12; and 12–14, 18); (*c*) 8/8 correct ($p < .01$) in comparisons in which enclosure was the only [illegible]plicable predictor (1–5; 2–4, 13; 4–11; 6–8; 10–14; 11–13; and 15–17); (*d*) 6/6 correct ($p < .05$) in comparisons in which orientation was the only applicable predictor (1–10; 3–12; 5–14; 6–15; 8–17; and 9–18); and (*e*) 2/2 correct ($p > .05$) in comparisons in which orientation and enclosure made compatible predictions (1–14 and 6–17).

A further breakdown of the 30 equivocal predictions shows (*a*) 19/20 correct (p [illegible] .001) when rectilinearity contradicted [illegible]tation but when rectilinearity domi[illegible] because one of the target arms was at 180° (prediction confirmed: 2–12, 18; 4–12, 18; 5–12; 6–7, 16; 7–8, 10, 14; 8–16; 10–16; 11–12, 18; 12–13; 13–18; 14–16; 15–18; and 17–18; prediction disconfirmed: 1–12); (*b*) 9/10 correct ($p < .01$) when orientation dominated the impotent 90°–135° rectilinearity difference (prediction confirmed: 1–2, 4, 11, 13; 4–5; 5–11, 13; 7–9; and 9–16; prediction disconfirmed: 2–5).

Although the above listing of predictions gives a precise and complete account of what the algorithm does and does not accomplish, it may be helpful to note a few of the key effects as they appear in Figure 2. First, the difference between solid and dashed functions reflects the effect of orientation helping Arm a in the first case and helping Arm b in the second (with the

exception of Patterns 2–11, 4–13, and 7–16, as discussed above). Second, enclosure accounts for the means to the left of the two 180° positions on the abscissa being higher than the corresponding means to the right. Thus, 1 > 5, 2 > 4, 10 > 14, 11 > 13, 6 > 8, and 15 > 17. Third, if Arm a's approach to rectilinearity facilitates grouping, why is there a drop in the function from Pattern 1 to 2 or from 5 to 4? The answer is that the 90°–135° step toward rectilinearity is the one that fails to enhance grouping, and that 2a and 4a lose the help of orientation that boosted 1a and 5a. Fourth, the rise from 14 to 13 and the incorrectly predicted rise from 10 to 11 were based on Arm a's gaining strength from orientation, the rectilinearity difference again being impotent. Fifth, the ordering 7b > 9b >3b (recall that when a target arm is a b arm, the sign of the mean rating must be reversed) can be explained as follows: 7b > 9b (orientation, rectilinearity impotent); 9b > 3b (rectilinearity); and 7b > 3b (rectilinearity and orientation). Sixth, the ordering 18b > 16b > 12b can be partially explained as follows: 18b > 16b (orientation, rectilinearity impotent); 18b > 12b (rectilinearity); 16b–12b, however, generates no prediction (a moderate rectilinearity effect favors 16b but orientation favors 12b). Seventh, 15b > 11b and 17b > 13b (rectilinearity and orientation). These, then, are a few of the bigger effects that catch the eye in scanning Figure 2.

Discussion

Overall, the simple algorithm employed above made 16/16 correct within-pattern predictions and 103/109 correct between-pattern predictions. It does not seem too much to say, then, that the algorithm generates an unprecedented bounty of predictions and that these turn out to be mostly correct.

The success of the algorithm suggests that Max Wertheimer's (1958) original observation that juxtaposition at 180° produces stronger grouping than at other angles needs to be elaborated in two ways. First, the function relating strength of grouping to angle of juxtaposition appears to be nonlinear—compared to the slope between 45° and 90°, the slope is relatively shallow between 90° and 135° and is relatively steep between 135° and 180°. Second, good continuation is not a function merely of the degree of rectilinearity; it is a compound of three phenomena—rectilinearity, orientation, and enclosure. As one of these can enhance or cancel another, any attempt to examine one must be guided by an awareness of all three.

References

Wertheimer, M. Untersuchungen zur Lehre von der Gestalt, II. (Trans. and abridged by Michael Wertheimer: Principles of perceptual organization). In D. C. Beardslee & Michael Wertheimer (Eds.), *Readings in perception.* Princeton, N. J.: D. Van Nostrand, 1958. (Originally published: *Psychologische Forschung*, 1923, 4, 301–350)

(Received July 9, 1973)

Journal of Experimental Psychology
1974, Vol 102, No. 5, 778–787

RECALL AND RECOGNITION OF SEMANTICALLY ENCODED WORDS [1]

ENDEL TULVING [2]

Yale University
and
University of Toronto, Toronto, Ontario, Canada

The hypothesis was tested that the phenomenon of recognition failure of recallable words is attributable to the discrepancy between semantic properties of encoded target words and the semantic interpretation of corresponding recognition test words. Target words were presented for study and encoding in the presence of specific list cues, and subsequently tested with strong associates of target words serving as extralist cues. In one critical encoding condition, the list cues were semantically congruous with target words; in another, they were identical with target words. Results showed that congruous encoding rendered extralist cues only slightly less ineffective than incongruous encoding, and that under the conditions of "identical" encoding extralist cues were least effective. The results thus provided little support for the hypothesis, suggesting instead that specificity of encoding of word events in episodic memory transcends the semantic meaning of words.

This article describes two experiments designed to explore the phenomenon of recognition failure of recallable words recently reported by Tulving and Thomson (1973). In the Tulving and Thomson experiments, target words T (e.g., CHAIR) were presented at input in the company of cue words C (e.g., *glue*). Subjects expected to be tested for recall of T with C as a cue. When, instead of presenting C as a cue, the experimenter provided another word X (e.g., *table*)—which was a close semantic associate of the target word T but which had not appeared anywhere in the list—as an extralist retrieval cue, subjects could not readily use it as an aid in recall of T. Retrieval of the target word T was rather poor even in a subject-generated recognition test. In this test, subjects were instructed to write down free-association responses to the cue X, and then were asked to identify the generated responses as copies of target words from the studied list. Subjects had no difficulty generating copies of target words, but they did have difficulty recognizing them as "old" words from the studied list. In one experiment, for instance, only 24% of generated target words were correctly recognized in such a subject-generated recognition test, although 63% of all target words were subsequently recalled when the word C that had been paired with T at input was presented as cue.

The effectiveness of extralist retrieval cues consisting of strong associates of target words is a well-established phenomenon, under a variety of experimental conditions (e.g., Bahrick, 1969; Light, 1972; McLeod, Williams, & Broadbent, 1971; Thomson & Tulving, 1970), and hence the lack of their potency under the conditions of the Tulving and Thomson (1973) experiments is of some interest. What renders extralist cues ineffective? Why cannot certain recallable words be recognized?

A general interpretation of all acts of successful and unsuccessful retrieval of information from episodic memory is given by the encoding specificity principle (Tulving & Thomson, 1973). It states that the properties of the memory trace of a word event are determined by specific encoding

[1] This research was supported by National Science Foundation Grant 24171X. The article was written while the author was in residence as a Fellow at the Center for Advanced Study in the Behavioral Sciences in Stanford, California. Nora Keane helped with experimental work, Mary Pori with statistical analysis.

[2] Requests for reprints should be sent to E. Tulving, Department of Psychology, Yale University, New Haven, Connecticut 06510, or at the Department of Psychology, University of Toronto, Toronto M5S 1A1, Canada.

operations performed on the input stimuli, and that it is these properties, rather than the properties of the word in semantic memory, that determine the effectiveness of any given stimulus as a retrieval cue for the event. The principle suggests that if a stimulus in the retrieval environment renders possible or facilitates recall of the target word T, the retrieval information was appropriate to or compatible with the information contained in the episodic trace of T. Conversely, if a particular stimulus is ineffective in retrieving a particular trace, the conclusion follows that the appropriate relation was lacking. It is not yet known exactly what constitutes the "appropriate" relation between information contained in the trace and information present in the general retrieval environment or in a specific cue. Nor is the nature of the retrieval process as such clear. In absence of such knowledge, we assume that the unknown relation is one of "similarity" or "informational overlap," and that the retrieval process consists of some sort of combining of information from the two sources.

The failure of recognition of recallable target words in the Tulving and Thomson (1973) experiments thus can be interpreted as a special case of encoding specificity. It can be assumed that the trace of the target word T that had been encoded in relation to a particular list cue C did not possess sufficient overlap with the copy of the target word that was generated as a semantic associate of the extralist cue X. This is why X failed to bring about recovery of information in the trace of the target T. The encoding specificity principle, however, says nothing further about conditions or mechanisms responsible for inadequate overlap or complementarity between the trace of the C–T complex and the copy of T. But, if one accepts the reasoning so far, it is possible to ask the next question: Why do the encoding operations performed by the system on the target word T sometimes create a trace whose informational content cannot be matched by that of cue X, a strong semantic associate of T? Thus, for instance, why does the trace of the target word CHAIR, when presented as a part of the *glue*–CHAIR compound, differ from the trace of the target word CHAIR presented by itself, so that in the former case cues *table* and *chair* are relatively ineffective, although they both are quite effective under different experimental conditions (Bahrick, 1969; Light, 1972; McLeod et al., 1971; Thomson & Tulving, 1970; Tulving & Thomson, 1973)?

One hypothetical answer to this question was studied and evaluated in two experiments reported in this article. The answer consists of two parts: (*a*) what is stored, or at least what is retained after an interval, about a target word in a typical list-item memory experiment is the "meaning" of the word, and (*b*) pairing of the target word T with a certain list cue C at input produces a specific meaning of T that differs from the meaning of the same lexical unit perceived or produced in a different context, such as in a free-association test in which some other word X elicits T. Thus, CHAIR presented in a to-be-remembered list of words as a part of the *glue*–CHAIR compound is assumed to have a different semantic representation in memory than CHAIR generated as a semantic associate to the word *table*. The failure of recognition of recallable target words, the argument goes, then comes about for the same reason as the decrement in recognition of a homographic target word when the semantic interpretation of its copy at the time of the test is changed through changes in its verbal context (e.g., Light & Carter-Sobell, 1970; Winograd & Conn, 1971).

This semantic explanation of encoding specificity manifested in the phenomenon of recognition failure of recallable words was tested in the experiments described here. The experimental paradigm used was one in which the lexical identity of the target words was held constant and in which the effectiveness of various retrieval cues was examined as a function of semantic overlap between encoding context of the target word and the retrieval cue. Thus, retrieval of target word T was tested with extralist cue X following the presentation of the target word in one of three input contexts: (*a*) in the company of cue

words whose dominant semantic meaning was relatively incongruous with that of the target word T and the extralist cue X; (*b*) with cue words whose meaning was congruous with that of the semantic target word and the extralist cue; and (*c*) with cue words that were identical with target words. Thus, for instance, with the word CHAIR serving as target, the Incongruous encoding context or condition was represented by the pair *glue*–CHAIR, the Congruous condition by the pair *furnish*–CHAIR, and the Identical condition by *chair*–CHAIR. The extralist retrieval cue was *table*. It was assumed that the semantic information contained in the retrieval cue *table* would be most compatible with the stored information about the target word CHAIR presented in the *chair*–CHAIR compound (Identical encoding), somewhat less compatible with CHAIR presented in the *furnish*–CHAIR compound (Congruous encoding), and least compatible with information about CHAIR that had appeared as a part of the *glue*–CHAIR compound (Incongruous encoding). In keeping with the semantic encoding hypothesis it was expected that the effectiveness of the extralist cue would be directly related to the semantic similarity or compatibility between the cue and the target word.

Experiment I

Method

Materials and design. Target words presented in three different encoding contexts were recalled in three successive tests. The three encoding conditions were all represented within a single list that contained 24 target words. Each target word was accompanied at presentation by another word, a list, or input cue. Three types of list cues defined the three encoding conditions: (*a*) incongruous, (*b*) congruous, and (*c*) identical. The designations refer to the semantic relations between target words or subsequently presented extralist cues on the one hand and the list cues on the other. Each type of list cue appeared with 8 target words in the list.

Three successive retention tests were those used in previous experiments by Tulving and Thomson (1973): (*a*) a test in which strong extralist associates of target words served as retrieval cues, (*b*) a recognition test in which copies of target words were generated by *S*s as free associations to strong extralist associates of target words, and (*c*) a test in which list cues served as retrieval cues.

The basic set of materials used in the construction of experimental lists is presented in Table 1. Each of the 24 target words is shown together with its extralist associate which was used in the first cued recall test, and with both the Congruous and the Incongruous input cue. The 24 target words were specifically selected from the materials used in earlier experiments described by Tulving and Thomson (1973) to yield low recognition scores. Extralist cues and the incongruous input cues of these 24 target words were the same as those used in earlier experiments. The identical list cues are not shown in Table 1, for obvious reasons. Extralist and incongruous cues had been selected in the earlier experiments (Thomson & Tulving, 1970) from two sets of free-association norms. In these norms target words were primary associates of extralist cues; they were given as responses to incongruous cues 1% of the time, and extralist and incongruous cues were not related. The congruous input cues were generated by *E*, for the purpose of the present experiment, to represent words closely related semantically to both the target words and their extralist associates. By perusing the listing of the experimental materials displayed in Table 1, the reader can convince himself that the congruous list cues are more closely related to the target word and the corresponding extralist cue words than are the incongruous cues.

For the purpose of balancing of specific words and their encoding conditions, three different lists were constructed of the materials that appear in Table 1. The 24 target words were divided into three subsets of 8, each subset being used with a different set of input cues (incongruous, congruous, and identical) in one of the three lists. The lists were used with equal frequency in the experiment (specifically, since there were 50 *S*s, with 17, 17, and 16 *S*s respectively, in each group). Half of the target words in each subset of 8 were tested in the first retention test (with extralist associates of targets as cues), while the other half was tested in the second test (subject-generated recognition test), with the two halves counterbalanced across the two tests. Thus, each of the 24 target words was represented with equal frequency in all conditions of the experiment.

Prior to the presentation of the critical experimental list, *S*s were given a single practice and set-establishing list. It consisted of 24 target words, half of which were accompanied by input cues identical with target words and half accompanied by incongruous cue words comparable to those used in the critical experimental lists.

Subjects and procedure. Fifty *S*s, undergraduate students at Yale University and the Southern Connecticut State College, served in Experiment I, either in fulfillment of course requirements or for pay. They were tested in small groups of from 2 to 4 persons, in a session lasting approximately 40 min.

The initial instructions, given prior to the presentation of the set-establishing list, informed *S*s of the general nature of their task and the type of materials to be presented. They were told that they

would see pairs of words presented on the screen in front of the room, that their task was to remember the capitalized target words, but that they should also pay attention to the cue word accompanying each target word and notice the relation between them, since these cue words would help them to remember the target words. No special mention was made of the fact that some of the pairs would consist of identical words, nor were *S*s told anything particular about the retention tests.

The cue–target pairs from the set-establishing list were presented by means of a slide projector, at the rate of 3 sec/pair. At the end of the presentation *S*s were asked to open their recall booklets that had been handed out to them at the beginning of the session, turn to page 3 in the booklet, read the instructions, and proceed with the retention test. On this recall sheet the 24 input cues from the list were presented in an order unrelated to the input order. The instructions at the top of the recall sheet told *S*s to write down the capitalized words that had accompanied the given word cues in the list. Subjects were given 3 min. for the completion of this test.

Next, *S*s were told that they would be shown a new list of pairs of words, and that their task was the same as before—to remember each capitalized target word, noting the relation between it and its cue word. The nature of the relation between target and cue words was not specified or further commented upon.

The 24 pairs of the experimental list were then presented, again at the rate of 3 sec/pair. At the end of the presentation *S*s were asked to turn to page 6 of the recall booklet, read the instructions on the top of the page, and then proceed. (The pages for recall tests in the recall booklet were always separated by two numbered blank pages, hence this numbering system.) The recall instructions informed *S*s that their task was to try to recall the capitalized words they had seen in the list, but that this time the cues presented on the page were different from those they had seen in the input list. The *S*s were told to look at each cue word given on the recall sheet, see whether it reminded them of any target words they had just seen, and, if so, write it down beside the cue word. Twelve extralist cues, corresponding to four target words from each of the three encoding conditions, appeared on the recall sheet. Subjects were given 3 min. for this task.

Next, *S*s were asked to turn to page 9 in the recall booklet, which contained 12 extralist cues corresponding to the target words that had not yet been tested. Extralist cues, each followed by four spaces for free-association responses, were listed in a column on the left-hand side of the page. The instructions at the top of the page told *S*s that their task now was to produce free associations to the presented cue words. They were to write down four words in response to each of the cue words they saw on the page, words that the cue word "made them think of." This was an unpaced task, and *S*s were given as much time as they needed to complete it. When all *S*s in the group had finished writing, they were told to look at each word they had written down, decide whether or not it had occurred as one of the target (capitalized) words on the preceding study list, and, if they thought it had, draw a circle around the word. Four minutes were given for the completion of this task.

Finally, *S*s were asked to turn to page 12 of the recall booklet, read the instructions on the top of it, and proceed as instructed. This page contained all 24 input cues, in a scrambled order with respect to the order of their original appearance in the study list. The instructions informed *S*s that these were the input cues from the list, and that their task was to write down as many target (capitalized) words from the study list as they could, each one beside its corresponding cue word. Four minutes were allowed for this task.

The experimental session concluded with *E* explaining the purpose of the experiment to *S*s and answering any questions they had.

TABLE 1

TARGET WORDS AND CUES

Target words	Extralist cues	Congruous cues	Incongruous cues
BALL	tennis	player	whistle
BLUE	sky	heaven	pretty
CHAIR	table	furnish	glue
COAT	lining	cloth	covering
COLD	hot	fire	ground
DAY	night	evening	sun
DIRTY	clean	wash	barn
FLOWER	bloom	leaf	fruit
FOOD	eat	meal	moth
GREEN	grass	lawn	cheese
HAND	finger	arm	tool
HARD	soft	smooth	glass
HIGH	low	down	hope
LARGE	small	size	stomach
LIGHT	dark	black	head
MAN	woman	child	command
NEED	want	desire	bath
OPEN	closed	shut	country
ROUND	square	circle	cabbage
SHORT	long	thin	stem
SLOW	fast	speed	memory
WATER	lake	fish	whisky
WET	dry	moisture	cave
WIND	blow	move	noise

EXPERIMENT II

The *S*s in Experiment I seemed to do considerably better in the first retention test (cuing with extralist associates) than in the second (subject-generated recognition test). Since these data suggested that the order of the two tests may have been an important source of variance in the retention scores, Experiment II was con-

ducted to replicate Experiment I in every respect, with the sole exception of the order of the two tests. In Experiment II, the first of three tests was the subject-generated recognition test, while the second one was the cued recall test with extralist associates as cues. Otherwise the procedure was identical with that used in Experiment I in all details. The same materials were used both in the set-establishing list and the experimental list, the same counterbalancing procedure was followed, and the same final test with input cues was given. Thirty-six new *S*s from the same sources as in Experiment I served in this experiment.

Results

Set-establishing lists. The mean number of words recalled from the first set-establishing list was 16.7 in Experiment I, and 17.8 in Experiment II. In both experiments slightly more target words were recalled to identical than to incongruous cues (M = 8.8 and 9.5 in the two experiments, respectively). The mean number of intrusion errors was 1.34 and 1.83 per S in Experiments I and II, respectively. Of these intrusion errors in the two experiments, 44% and 58%, respectively, were "identical" intrusions: S gave a response identical with the cue although in the input list the cue had been paired with another word. It is possible that the higher recall of target words to identical cues was at least partly a consequence of inflation of recall scores by unidentifiable intrusions, but since it is not entirely clear what would constitute an appropriate correction, no attempt was made to assess the possible guessing bias.

Scoring and analysis. Analysis of the cued recall data, in all tests in which extralist or input cues were used, was straightforward. Each S's recall score was the number of target words recalled, from a maximum of four or a maximum of eight per condition in the two tests, respectively. A stringent scoring criterion was used: S received credit for every target word only if it was correctly paired with its cue. (The inclusion in the data of target words recalled to incorrect cues, or to no particular cues, would not have changed the conclusions.)

Intrusion errors were ignored in all analyses. It is not immediately obvious how their inclusion in the results would change the overall conclusions drawn from the experiments.

In scoring the protocols from the subject-generated recognition tests, *S*s were given credit only for those copies of target words generated and recognized that belonged to the half of the input list being tested in the recognition test. In Experiment I there was a total of 33 cases in the protocols of the 50 *S*s where a copy of the target word had been generated as a part of the free-association procedure and correctly recognized even though these words nominally belonged to the other half of the list that had been tested in the immediately preceding cued recall test using extralist retrieval cues. Since it was impossible to rule out the possibility that the words from the other half of the list constituted importations from the immediately preceding test rather than representing *S*s' memory for the material seen in the input list (particularly since most of these generated and recognized words had been in fact recalled by *S*s in the immediately preceding test), they were excluded from the analysis.

The data from the subject-generated recognition tests required a special treatment inasmuch as the maximum possible score of correct identification of target items depended upon the number of copies of target items any particular S had generated in the free-association test. There were three essential steps in the analysis of these data.

First, the number of copies of target items correctly generated in the free-association test were tabulated for each S, separately for target words from each of the three encoding contexts. The means of these scores were 2.84, 2.76, and 2.91 for the incongruous, congruous, and identical encoding conditions, respectively, in Experiment I, and 2.98, 3.11, and 2.44, for the same three encoding conditions in Experiment II. There is no obvious explana-

tion for the lower value of this statistic in the identical encoding condition in Experiment II, and since the same decrement was not observed in Experiment I, whatever implications this finding might have will be ignored in what follows.

The second step consisted of the calculation of proportions of copies of target words that were recognized by each *S* in each encoding context. Thus, for instance, if a given *S* generated three copies of the target item in a particular encoding condition, and recognized one of these, his recognition score (hit rate) was $\frac{1}{3}$ or .33. The means of these proportions are provided as summary statistics for the subject-generated recognition test in Table 2. In a small number of cases, when *S* did not generate any copies of target words for a particular encoding condition in the free-association test, and hence could not possibly have recognized any, his score was considered indeterminate and was not entered into any analyses.

The third step in the analysis of the recognition scores was necessary for the purposes of the statistical evaluation of the data. Since the distributions of the recognition hit rates from individual *S*s in different encoding conditions in most cases were not distributed normally, statistical tests involving these measures were deemed inappropriate. Another measure of recognition performance was used instead. For each *S* in each encoding condition a right minus wrong (R − W) score was calculated.[3] Each target word whose copy had been generated and correctly recognized was considered R while each generated copy of a target word that was not recognized was considered W. For instance, an *S* who generated three copies of targets in a particular encoding condition, and recognized one of these, received an R − W score of $1 - 2 = -1$. The *differences* of these R − W scores for all three comparisons (incongruous vs. congruous, incongruous vs. identical, and congruous vs. identical) were distributed normally, thus permitting the use of *t* tests for related measures. Thus, while the data in Table 2 are summarized as mean hit rates, the corresponding analyses were conducted on the R − W scores as explained. There was, of course, a high although not a perfect correlation, within each of the encoding conditions, between individual *S*s' hit rates and R − W scores, the coefficient exceeding .90 in all cases.

Retrieval data. Table 2 presents a summary of the data on retrieval of target words from three different encoding conditions in three different tests.

Consider first the mean proportions of target words retrieved in various tests, with data pooled over the three encoding conditions in each test. Two observations are of interest. First, the order of the first two tests, both involving extralist cues, appears to have exerted considerable effect on the level of retrieval: *S*s' performance was higher in the first test (cued recall with extralist cues in Experiment I, and subject-generated recognition test in Experiment II) than in the second test. To evaluate test-order effects, a *t* test for unrelated measures was done comparing the differences between *S*s' scores on the cued recall and recognition tests in Experiment I with the same difference scores in Experiment II. The obtained value for t (84) was 4.57, $p < .01$. It appears, therefore, that some general test interference was produced in the course of the first test that suppressed the level of performance in the second. The proportions of copies of target items generated in the recognition test were identical at 71% in both experiments, suggesting that it is only the episodic and not the semantic (Tulving, 1972) component of the task that reflects interference, but given the design of the experiments, it is difficult to say anything more about the nature and sources of the test interference.

The second point with respect to retrieval scores pooled over different encoding conditions concerns the superiority of the performance in the final cued recall test over that in the other two tests, in both experiments. Statistical evaluation of the data was accomplished by means of *t* tests

[3] I am grateful to Perry Gluckman for suggesting this analysis.

TABLE 2

MEAN PERCENTAGES AND STANDARD DEVIATIONS OF TARGET WORDS RECALLED AND RECOGNIZED IN THREE TESTS AS A FUNCTION OF ENCODING CONTEXT

Experiment and encoding context	Test: Cued recall, extralist cues		Subject-generated recognition		Cued recall, list cues	
	M	*SD*	*M*	*SD*	*M*	*SD*
Experiment I ($N = 50$)						
Incongruous	51.5	34.0	26.3[a]	31.5	60.5	21.6
Congruous	52.0	27.1	36.5[a]	32.3	49.8	22.4
Identical	39.0	30.0	24.6[a]	31.1	46.5	26.1
M	47.5	23.4	28.6[a]	22.9	52.2	16.9
Experiment II ($N = 36$)						
Incongruous	28.5	29.4	42.4[b]	35.7	51.0	23.2
Congruous	36.1	29.6	46.1[b]	30.4	51.4	22.9
Identical	22.9	25.6	26.8[b]	27.2	36.8	30.3
M	29.2	20.1	39.9[b]	22.5	46.4	19.3

[a] $N = 45$.
[b] $N = 32$.

after the data in various tests were normalized in order to equate the three tests for maximum scores possible. The resultant t test was highly significant when the performance in the third test was compared with the performance in the second test in each of the two experiments, t (49) = 6.75 in Experiment I, and t (35) = 5.28 in Experiment II, both ps < .01. The t values were not significant when recall with input cues was compared with probability of retrieval in the first test in each of two experiments, t (49) = 1.74 in Experiment I, and t (35) = 1.88 in Experiment II, in both cases $.05 < p < .10$.

The data of primary interest concern differences in retrieval of target words that had been presented in different input contexts in the study list. The differences between encoding conditions were evaluated by means of t tests for related measures, separately for each of the three tests in both experiments. The results of these tests are tabulated in Table 3.

Although there was a slight numerical superiority in extralist cued recall and subject-generated recognition scores in the congruous input cue conditions over those in the incongruous input cue conditions, statistical analyses of the data, summarized in Table 3, showed that these differences did not reach acceptable degrees of reliability. The largest difference between the incongruous and congruous encoding conditions was observed in the subject-generated recognition test in Experiment I (26.3% vs. 36.5%), yielding a t (44) = 1.47, $.05 < p < .10$. In Experiment I, recall in response to list cues of target words that had appeared in the context of incongruous input cues was higher (60.5%) than recall of target words that had appeared in the context of congruous encoding cues (49.8%), t (49) = 2.80, $p < .01$, and this difference might be regarded as tempering the conclusion about lack of evidence for the effect of encoding conditions on retrievability of target words by extralist cues or their copies, but since Experiment II did not replicate this finding, its implications cannot be taken too seriously.

Since each to-be-remembered list word served as a target word in two successive tests, it was possible to examine recall or nonrecall of individual words in one test in relation to their recall or nonrecall in the other test. Indeed, the phenomenon of recognition failure of recallable words that provided the starting point for the present experiments can be most directly defined by demonstrating that subjects cannot

TABLE 3

SUMMARY OF t STATISTICS FROM TESTS COMPARING RECALL AND RECOGNITION SCORES BETWEEN PAIRS OF ENCODING CONDITIONS

Comparison	Cued recall, extralist cues		Subject-generated recognition		Cued recall, input cues	
	Experiment I	Experiment II	Experiment I	Experiment II	Experiment I	Experiment II
Congruous vs. Incongruous	.11	1.12	1.47	.78	−2.80**	.08
Incongruous vs. Identical	2.31*	1.10	.48	2.31*	3.10**	2.82**
Congruous vs. Identical	2.99**	2.45**	2.19*	2.68**	.88	2.60**

* $p < .05$.
** $p < .01$.

identify copies of target words as "old" but can produce these *same* words in presence of list cues. In the analysis of the "fate" of individual words that follows, retrieval of target words in extralist cue tests is compared with their retrieval in the final list-cue test.

Frequencies of words that are recalled in *both* the extralist cue test and the final list-cue test, as well as words that are recalled in *neither* of these two tests, are of little immediate interest. These data may reflect general "ease" or "difficulty" of words, breadth of encoding or failure of storage, effects of recoding or strengthening of traces in the first test, and other similar factors. Words that were retrieved only in the extralist cue test or only in the list-cue test provide more interesting data from the point of view of encoding specificity. These data are tabulated in Table 4.

Table 4 lists total frequencies of target words that were recalled or recognized in the extralist cue test but not recalled in the final list-cue test, and frequencies of target words not recalled or recognized in the extralist cue test but recalled in the final test. These two categories of words will be referred to as RN words and NR words, respectively, indicating the words' recall (R) or nonrecall (N) in the first (extralist cue) and the second (list cue) test. The left-hand panel of Table 4 presents data for target words initially tested in the extralist cued recall test, and the right-hand panel contains data for target words tested in the subject-generated recognition test.

Each entry in Table 4 consists of two figures. The first gives the total frequency (with the data pooled over all 50 *S*s in Experiment I and 36 *S*s in Experiment II) of RN words, while the second represents the total frequency of NR words. Each entry thus can be thought of as a ratio of two quantities, the number of RN words to the number of NR words, and the numerical value of this ratio, or some transformation of it, could be used as a rough index of the magnitude of the recognition failure of recallable words or the magnitude

TABLE 4

TOTAL FREQUENCIES OF TARGET WORDS RECALLED ONLY IN TEST WITH EXTRALIST CUES, IN RELATION TO FREQUENCIES OF WORDS RECALLED ONLY IN TEST WITH LIST CUES

Encoding condition	Recall with extralist cues			Subject-generated recognition test		
	Experiment I	Experiment II	Total	Experiment I	Experiment II	Total
Incongruous	17/45	15/54	32/99	10/47	17/26	27/73
Congruous	27/33	9/32	36/65	15/29	16/28	31/57
Identical	15/40	8/33	23/73	8/30	11/20	19/50
Total	59/118	32/119	91/237	33/106	44/74	77/180

of the failure of extralist cues in retrieval of words known to be represented in the episodic memory store.

Three features of the data in Table 4 should be mentioned. First, the data, pooled over both experiments, are rather similar in the two panels of Table 4, suggesting that underlying retrieval processes are probably not entirely dissimilar in extralist cued recall test and the subject-generated recognition test. Second, in all three encoding conditions, in both experiments and in both tests, the number of target words accessible only through list cues (NR words) was higher than that of words accessible only through extralist cues (RN words). Third, extralist cues were somewhat more effective for target words encoded in the congruous input condition than for those encoded under the other two conditions, as judged by the ratio of RN words to NR words, although the relatively small number of observations on which these data are based renders the differences statistically unreliable.

Discussion

The primary purpose of the experiments was to evaluate the hypothesis that specificity of encoding manifested in the phenomenon of recognition failure of recallable words (Tulving & Thomson, 1973) is a matter of specificity of semantic meaning of the to-be-remembered words and retrieval cues. Since in earlier experiments the target words were encoded in relation to list cues possessing little semantic similarity to target words and extralist cues, it seemed reasonable to argue that considerable semantic discrepancy existed between the target word as stored in that particular context and the literal copy of the target word presented as retrieval cue in the subject-generated recognition test. The graphemic units might have been the same in both cases, but they represented different bundles of semantic information.

The results of the present experiments do not provide much support for this semantic interpretation of encoding specificity. Some facilitation of retrieval of target words by extralist cues was observed under conditions where target words had been encoded in relation to cue words semantically much more compatible with both target words and extralist cues, in comparison with retrieval of target words encoded in relation to less congruous cues, but this facilitation was rather small and statistically not reliable. Furthermore, a considerable number of target words encoded in presence of congruous input cues were still not recognized in the subject-generated recognition test and could not be recalled to strongly associated extralist cues, even though they were recalled subsequently in presence of list cues, indicating that relevant information was available in the store. Finally, extralist cues were least effective, both in the recall and the subject-generated recognition tests, when target words were presented under the identical encoding condition where the input cue was nominally most compatible with the target word.

It is entirely possible that the hypothesis of semantic determination of encoding specificity fared so poorly in these experiments simply because the congruous encoding conditions were not congruous enough and the identical encoding conditions did not produce semantic encoding. If greater care were taken to ensure a high degree of semantic overlap between list cues and extralist cues the hypothesis might give a better account of itself. Nevertheless, the fact that it is not easy to eliminate the encoding specificity effect by manipulating the relation between input cues and extralist cues suggests that encoding of a familiar word in certain episodic memory tasks (Tulving, 1972) may entail degrees of specificity that cannot be achieved by changing semantic context of words. It may be more appropriate, therefore, to think about properties of memory traces of particular events and episodes as going beyond the properties of words as communicative units of language. It is [illegible] clear what these nonsemantic properties are.

The low level of recall and recognition of target words that had been presented in the input list under identical encoding conditions was quite unexpected. Several previous experiments, cited in the introduction, have shown that associated extralist retrieval cues are quite effective in facilitating recall of target words when these words occur in the study list without any manipulated context, and when the subjects are left free to encode them any way they want. There was no particular reason to expect, before doing the experiments described here, that the presentation of two identical copies of a word, one as "cue" and the other one as "target," would make any

substantial difference in the encoding of these list words. Yet the data were quite clear in showing that the Identical encoding condition was associated with lowest levels of recall and recognition, both in extralist cue tests and the final input cue test. It is this finding that appears to be most difficult to reconcile with the "semantic" interpretation of encoding specificity, suggesting that other or at least additional factors are involved.

The data from the practice list, in which immediate recall of target words was just about the same under identical encoding and incongruous encoding conditions, suggest that it was not the initial level of storage of "identically" encoded words that was impaired but rather that the low level of subsequent recall was some consequence of events that took place after the study of the list. It is difficult to say anything definite about this impairment on the basis of existing data. It may be that subjects encoded target words under identical conditions primarily in terms of their phonetic properties, in terms of a clang association between the cue and the target word, and that this information was lost more rapidly than the specific "semantic" information, or that the subsequently presented retrieval cues were somewhat less appropriate for providing access to this kind of information. It is also conceivable that the low level of retrieval of target words under identical conditions was a consequence of a high degree of similarity of encoding operations performed on these target words during the presentation of the list. If part of the information stored in the memory trace of an event represents the encoding operation performed on the input (Tulving & Thomson, 1973), then a high degree of similarity of encoding operations performed on a number of list items could render the resulting traces less unique and hence somehow less readily retrievable (Lesgold & Goldman, 1973; von Restorff, 1933). Be it as it may, the present data rather clearly suggest that the identity relation between two items, one designated as cue and the other as target, does not represent the extreme position on the dimension of associative or semantic relatedness.

Finally, in relation to earlier experiments (Tulving & Thomson, 1973) the present data showed rather a high degree of effectiveness of extralist retrieval cues, despite the fact that the materials for the present experiment were specially selected from among those used in earlier studies to yield a minimal level of retrieval in presence of extralist cues. It was hoped that such minimal levels of retrieval would have made it easier to demonstrate the effect of a congruous semantic input context. It may be that the use of a study list containing different sorts of cue–target relations was responsible for the discrepancy between the present and earlier experiments with respect to effectiveness of extralist cues, although for the time being this must remain pure speculation.

REFERENCES

Bahrick, H. P. Measurement of memory by prompted recall. *Journal of Experimental Psychology*, 1969, **79**, 213–219.

Lesgold, A. M., & Goldman, S. R. Encoding uniqueness and the imagery mnemonic in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 193–202.

Light, L. L. Homonyms and synonyms as retrieval cues. *Journal of Experimental Psychology*, 1972, **96**, 255–262.

Light, L. L., & Carter-Sobell, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 1–11.

McLeod, P. D., Williams, C. E., & Broadbent, D. E. Free recall with assistance from one and from two retrieval cues. *British Journal of Psychology*, 1971, **62**, 59–65.

Thomson, D. M., & Tulving, E. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 1970, **86**, 255–262.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

Tulving, E., & Thomson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352–373.

Von Restorff, H. Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung*, 1933, **18**, 299–342.

Winograd, E., & Conn, C. P. Evidence from recognition memory for specific encoding of unmodified homographs. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 702–706.

(Received July 9, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 788–794

BODY TEMPERATURE AND TEMPORAL ACUITY [1]

JAMES F. O'HANLON,[2] JAMES J. McGRATH,[3] AND MICHAEL E. McCAULEY

Human Factors Research, Inc., Goleta, California

Two experiments were conducted to determine the relationship between temporal acuity and body core temperature; a third experiment served as a control. In Experiment I, thresholds for the perception of the succession of dichoptic light flashes were tested when body core temperature was normal and after it was increased to 38.3°C. by diathermy. In Experiment II, critical flicker frequency (CFF) thresholds and an electroencephalogram (EEG) were measured during euthermic, hyperthermic (39.0°C.), and recovery conditions. Results of both experiments indicated that temporal acuity increased with hyperthermia. Perception of succession and CFF were both facilitated by increasing body core temperature. The EEG alpha frequency also varied directly with body temperature, but was not significantly correlated with CFF. In Experiment III, no change in pupil diameter was found during hyperthermia, eliminating an interpretation of the results of Experiments I and II based on the Ferry-Porter law. These results were interpreted as supporting "psychological moment" theory.

A common premise shared by many theorists (Craik, 1948; Kristofferson, 1967; McReynolds, 1953; Shallice, 1964; Stroud, 1955; Walter, 1950; White, 1963; Wiener, 1948) is that the nervous system does not process all sensory inputs continuously, but briefly samples each input and routes selected inputs serially into the perceptual process. The time required to complete one sampling cycle has become known as the "psychological moment." According to Harter (1967), an assumption frequently made by the "moment" theorists is that the nervous system codes all sensory information in terms of sequential order of occurrence. All information arriving in the nervous system within one moment is encoded similarly and perceived as instantaneous and simultaneous; information arriving over the course of more than one moment is perceived as having duration. Discrete stimuli which are perceived in different moments are perceived as differing in time of occurrence.

According to Stroud (1955, p. 181), identical visual events that differ in time of occurrence are psychologically differentiated at a maximum rate of one per moment. By his reasoning, temporal acuity (the ability to discriminate between the occurrence of events in time) would be proportional to the moment frequency.

Still other theorists have used the moment concept to explain the perception of time passing (Treisman, 1963; White, 1963). Their premise has been that somehow the nervous system registers the accumulation of moments between events in time in a manner permitting the perception of the relative durations between those events. To those authors, the mechanism responsible for the moment also functions as a neural pacemaker which governs the perceived rate of time passing. That assumption implies that a factor which influences the neural pacemaker would affect both temporal acuity and the perception of time passing. Whether hyperthermia is such a factor is the question addressed in the present research.

Induced hyperthermia has been found to increase the perceived rate of time passing; i.e., subjective estimates of a unit of time are systematically shortened as body temperature increases (Hoagland, 1933; Kleber, Lhamon, & Goldstone, 1963; O'Hanlon, Danish, & McGrath, 1968). Hoagland attributed this result to the physiochemical effect of temperature on a hypothetical

[1] Supported under Contract DA-49-193-MD-2743 by the U. S. Army Medical Research and Development Command.

[2] Requests for reprints should be sent to James F. O'Hanlon, Human Factors Research, Inc., Goleta, California 93017.

[3] Now at Anacapa Sciences, Inc., Santa Barbara, California.

pacemaker located in the central nervous system. However, some investigators (Bell, 1965; Bell & Provins, 1963) have failed to find any significant relationship between hyperthermia and performance in time-keeping tasks. But those negative results have been attributed to possible confounding effects of interpolated tasks and the relatively small induced changes in body temperature (Fox, Bradbury, Hampton, & Legg, 1967).

Various investigators have sought to determine whether induced hypothermia, as well as hyperthermia, affects time perception in the manner predicted by Hoagland's pacemaker model. In two studies (Fox et al., 1967; Lockhart, 1971), an increase in the perceived rate of time passing was found in both hyperthermic and hypothermic conditions. However, Baddeley (1966) found a significant decrease in the perceived rate of time passing in scuba divers exposed to cold water as predicted by the model. Further support was provided by O'Hanlon et al., who found that the perceived rate of time passing decreased with hypothermia and increased with hyperthermia in the same *S*s. Therefore, while the evidence is by no means unanimous, there appears to be a substantial degree of support for the existence of a temperature-dependent neural pacemaker.

If the neural pacemaker and the psychological moment are related, as suggested by White (1963), then hyperthermia would be expected to increase temporal acuity. Thus, the major purpose of the experiments described here was to determine if temporal acuity, like the perception of time passing, varies with hyperthermia. A secondary purpose of this research was to ascertain whether a change in the brain's dominant alpha rhythm parallels any hyperthermia-induced change in temporal acuity. Several authors have suggested a relationship between alpha frequency and temporal acuity (Harter, 1967; Kristofferson, 1966, 1967), but no conclusive evidence has been offered (Legg, 1968).

Two separate experiments were performed to determine the relationship between temporal acuity and hyperthermia. In the first, temporal acuity was determined from a test of the perception of succession (Type 2 test, Fraisse, 1963); in the second, it was determined using a version of the test for critical flicker frequency (CFF). Electroencephalogram (EEG) recordings were taken during the second experiment. A third experiment was conducted to ascertain whether hyperthermia changes pupillary diameter in any way which might affect interpretation of the results from the former experiments.

Experiment I

Method

Subjects. The *S*s were 11 male 21–28-yr.-old college students, who were paid for their participation. All had experienced induced hyperthermia in a previous experiment.

Experimental setting. The laboratory consisted of two adjacent rooms. The *S*'s room contained a standard hospital bed, a shortwave diathermy machine (Liebel-Flarsheim, No. SW-660), and a visual display. The *E*'s room contained apparatus for controlling the display and recording *S*s' temperature and responses.

Apparatus. The visual display was a standard Lafayette binocular viewing chamber (No. 1202C). The display contained two Sylvania glow-modulator lamps (No. R1166), each of which illuminated a circular diffusion screen with a diameter of 1.3 cm. The two diffusion screens were 7.0 cm. apart and were viewed from a distance of 38 cm. through an eyepiece. An opaque septum within the chamber permitted dichoptic stimulus presentation. Each lamp provided a flash luminance of about 41.56 cd/m² against a background of less than 16 cd/m² at the eyepiece. The display control was a modified Lafayette flicker-fusion unit (No. 1202A). The modification permitted the administration of two flashes, 10 msec. in duration, one to each eye. Two Hunter timers controlled the duration of the interval between flashes of the two lights. Tympanic temperature, which approximates brain temperature (Benzinger & Taylor, 1963) was recorded by a Digitec digital thermometer (No. 500) connected by cable to a thermistor fixed in *S*'s external auditory meatus, against the tympanic membrane.

Task. Two pairs of light flashes were presented in the visual display; one pair was simultaneous, the other was successive. The onsets of successive flashes were separated by 10-, 20-, 30-, or 40-msec. intervals, and all flashes were of 10-msec. duration. Upon observing both pairs of flashes, *S* was required to state which pair consisted of successive flashes. If he was unsure, both pairs were presented again. The task conformed to the psychophysical method of constant stimuli with a forced-choice response.

Procedure. Each *S* was scheduled for a sequence of three sessions with 1 wk. between sessions.

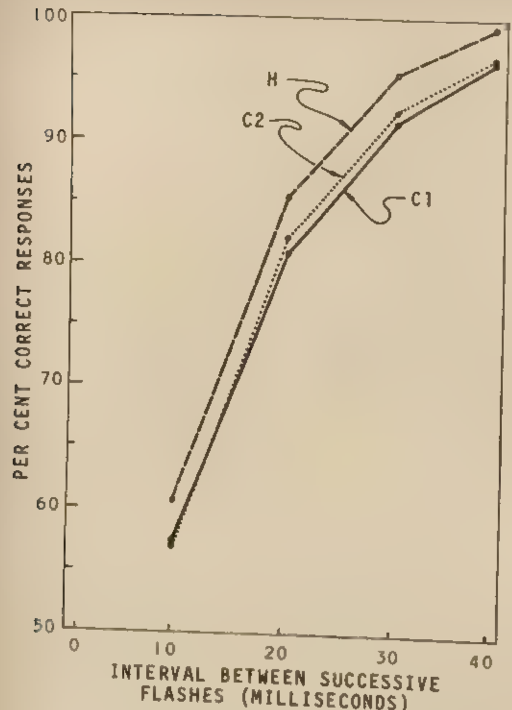


FIGURE 1. The percentage of correct responses by 11 *S*s as a function of the duration between successive flashes in the first control condition (C_1), hyperthermia condition (H), and the second control condition (C_2).

The first session was a training condition. The second session was comprised of a control condition (C_1) followed by the hyperthermia condition (H). The second control condition (C_2) was administered during the third session. In each session *S* wore a sweat suit and reclined on the bed under blankets. Upon request, he was given cold water from a thermos, which was replenished whenever necessary. His tympanic temperature was recorded continuously.

In the training condition, *S* performed the task in 40 trials with knowledge of results after each trial, and then in four blocks of 40 trials without knowledge of results. In C_1, *S* performed the task

TABLE 1

SUMMARY OF A NEWMAN–KEULS TEST OF DIFFERENCE IN TOTAL NUMBER OF CORRECT RESPONSES BY 11 *S*, COMPARING ALL POSSIBLE PAIRS OF CONDITIONS

Comparison	Total difference	*q*	*df*	*p*
H–C_1	63	6.80	3, 20	<.01
H–C_2	55	5.94	2, 20	<.01
C_1–C_2	8	.86	2, 20	*ns*

Note. Abbreviations: H = hyperthermia, C_1 = first control condition, and C_2 = second control condition.

in four blocks of 40 trials. Between [illegible] he rested for 5 min. At the end of C_1, *S* reste[illegible]r 15 min., and then began H. In H, diathermy [illegible] applied to the thorax, abdomen, and thighs, [illegible]hat *S*'s tympanic temperature was raised from [illegible]erage euthermic level of 37.0°C. to 38.3°C. [illegible]om 60–90 min. of diathermy were required to [illegible]omplish this change in body temperature. [illegible] the criterion temperature, diathermy was discont[illegible]ed, and *S* performed the task in four blocks of 40 [illegible]als each. Tympanic temperature was maintaine[illegible] between 38.25°C. and 38.35°C. by applying diath[illegible]y whenever necessary during the 5-min. rest per[illegible]ds between blocks of trials. In C_2 the procedure [illegible] the same as in Condition C_1.

In all conditions, each block of 40 trials had [illegible] following characteristics: (*a*) In 20 trials, the pa[illegible] of flashes were simultaneous–successive, and in [illegible] trials they were successive–simultaneous. (*b*) Th[illegible] sequence of flashes in the successive pair was left–right in 20 trials, and right–left in 20 trials. (*c*) Each interflash interval (10, 20, 30, or 40 msec.) occurred 10 times in the 40 successive pairs. Each interflash interval was used an equal number of times with each order of pairs and each sequence of flashes. The order of pair presentations and the order of interflash intervals were randomized independently within each block of trials.

Results

Training session data from each *S* were inspected to ensure that he had arrived at a stable level of performance. These data were not analyzed further.

The percentage of correct responses was tabulated for each interflash interval in Conditions C_1, H, and C_2. These percentages are shown in Figure 1 as functions of the duration of the interflash interval. The percentage of correct responses was greater in H than in either of the control conditions for all values of interflash intervals. The differences among performances in the C_1, H, and C_2 conditions were tested for statistical significance using a two-way mixed-design analysis of variance of the total number of correct responses. That analysis revealed that the differences among total correct responses in the three conditions were significant, F (2, 20) = 3.74, $p < .05$. A Newman–Keuls test was applied to determine the significance of the differences between all possible pairs of conditions. The results are summarized in Table 1.

The difference in the total number of correct responses between the H session

and each control condition was significant beyond $p < .01$. There was no significant difference between the control conditions.

Experiment II

Method

Subjects. The same 11 men who had served as *Ss* in the first experiment served in the second.

Apparatus. The display control apparatus was modified for estimating *Ss'* CFF, according to the version of the psychophysical method of limits developed by von Békésy (1947). Briefly, the lamps were made to flash in unison with a 50/50 light to dark ratio, and the flash frequency was made to vary continuously at a rate of .5 Hz/sec. This was done with a reversible servomotor which drove the oscillator that controlled the flicker rate, as well as a potentiometer which yielded a voltage analog of flicker frequency to drive a pen in a strip-chart recorder. After initial adjustment, *S* determined the duration of frequency change by depressing a handheld switch which caused the servomotor to reverse direction after an 8–10-sec. delay. These responses were recorded by a brief pen deflection on the chart recording.

Two channels of a Gilson polygraph (M8PM) recorded EEG from Grass silver subdermal electrodes placed over *S*'s right and left occipital cortices. The active recording electrodes were referenced to linked, clip-type, silver electrodes attached to both earlobes.

Task. The task consisted of pressing the response switch whenever the lamp's brightness appeared constant after it had been flickering, or whenever the brightness seemed to flicker after it had been constant. The former response marked *S*'s fusion threshold, the latter response marked his flicker threshold. The midpoint of a line between fusion and flicker thresholds on the chart recording was the measurement of CFF.

Procedure. Each *S* participated in three sessions scheduled 1 wk. apart. Conditions C_1, H, and C_2 were administered in successive sessions.

In C_1 and C_2 the task was administered in four 10-min. tests, interspaced with 5-min. rests. Recordings of EEG were made during 5-min. periods before the first test and after the final test. During all EEG recordings, *S* reclined quietly with his eyes closed.

In H, *S* rested as before for 5 min. while his EEG was recorded. Next, he undertook a 10-min. version of the CFF test. Diathermy was then administered to *S* for 30 min., and his CFF was tested as before. The sequence of heating and testing was repeated until *S*'s tympanic temperature had risen by 2.0°C. This took 90–120 min. to complete. After reaching the criterion temperature, *S* was allowed to recover to his normal temperature. During recovery, *S*'s CFF and EEG were measured in alternate 10- and 5-min. periods. Recovery lasted 55–75 min.

Eight separate CFF determinations were made in each 10-min. test. These were averaged to provide a single CFF measurement for each *S* in each test.

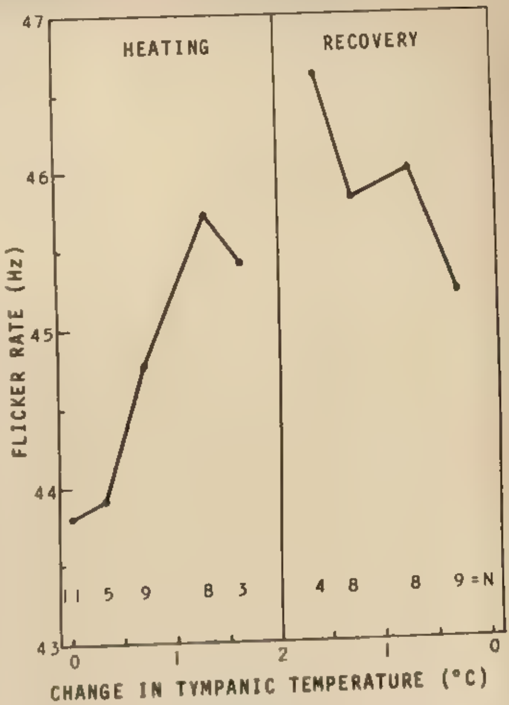


Figure 2. Mean critical flicker frequency as a function of the change in tympanic temperature during the hyperthermia condition.

Results

In both control conditions, C_1 and C_2, *Ss'* mean CFF varied in a nonsignificant manner between about 43.5 Hz. and 44.3 Hz. In H, mean CFF varied as a function of tympanic temperature as shown in Figure 2.

The data in Figure 2 were obtained by computing the mean CFF for each .5°-C. interval of changing tympanic temperature. However, all *Ss* were not tested within the same .5°-C. intervals due to differential body heating rates, so the mean values shown in Figure 2 were necessarily based upon a different number of observations. The *N* for each mean is included in Figure 2 below the respective data point. The product-moment correlation coefficient between mean CFF and mean tympanic temperature was $r = .78$, $p < .05$.

The EEG recordings from one *S* were practically devoid of alpha activity. These data were not used. The mean ($\pm SE$) alpha frequency for the remaining 10 *Ss* was 10.7 Hz. ($\pm .25$) in both C_1 and C_2.

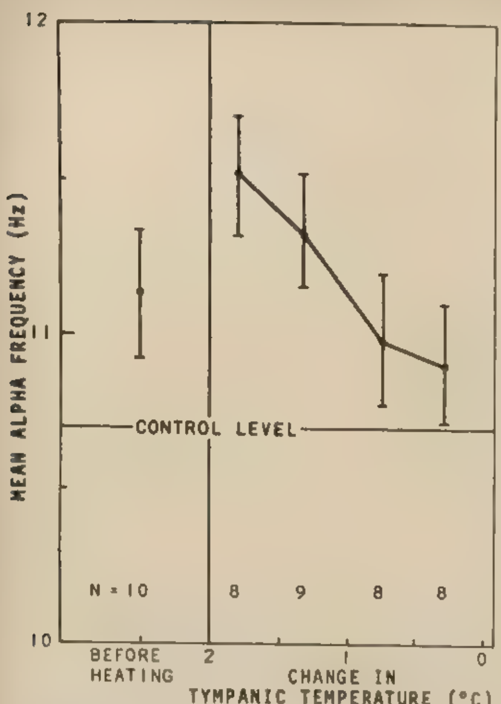


FIGURE 3. Mean alpha frequency ($\pm SE$) before hyperthermia and as a function of the change in tympanic temperature during recovery from hyperthermia.

During H, mean ($\pm SE$) alpha frequency varied with tympanic temperature as shown in Figure 3.

Even before heating, the mean alpha frequency was higher in H than the 10.7 Hz. of C_1 and C_2. There was a further increase upon heating the *Ss*, up to a maximum mean of 11.5 Hz., followed by a decline as their temperatures returned to normal. The difference between the mean alpha frequency in C_1 and the first measurement (preheating) in H approached significance, t (9) = 2.19, $p < .10$. The difference between mean alpha frequency in C_1 and the second measurement in H (i.e., during maximum hyperthermia) was significant, t (7) = 4.05, $p < .01$. However, the difference between the first and second measurement in H was not significant, t (7) = 1.27.

In spite of the increase in mean alpha frequency during heating, no significant correlation was found between alpha frequency and CFF in either C_1 or H. The correlation between the changes in CFF and in alpha frequency that occurred with changing tympanic temperature was similarly nonsignificant.

EXPERI[illegible]

Interpretation of th[illegible] from Experiment II was made di[illegible] the failure to control or measure [illegible]pillary diameters while their core tem[illegible]es changed. A systematic increase in [illegible] diameter with temperature *might* have [illegible]used the observed change in CFF [illegible] to the Ferry–Porter law (Brown, [illegible]). The relationship between core temperature and pupillary diameter was not revealed by an extensive review of the literature. Therefore Experiment III [illegible] supply the data requi[illegible]g whether any such rel[illegible]

Method

Subjects. The *Ss* were three 27–34-yr.-old healthy males.

Apparatus. The diather[illegible] used to induce hypertherm[illegible] complished with a Nikon F[illegible] Nikkor 55 mm. F3.5 len[illegible] was measured from photographic enlargements with dial calipers (Mitutoyo, No. 505-626). Rectal temperature was measured, using the Digitec thermometer.

Procedure. The *Ss* were heated to a temperature of 39.0°C. Heating was accomplished in 37–81 min., depending on the *S*. At 15-min. intervals during heating, a series of three photographs of *S*'s left eye were taken within a 30-sec. period. After reaching criterion temperature, *Ss* were allowed to recover, and a similar series of photographs was taken at approximately 15-min. intervals, until a drop of 1.0°C. during recovery. This degree of recovery was obtained in 16–40 min., depending on the *S*. During photography, *Ss* looked at a fixation point on a blank field at a distance of 38 cm. Luminance at *S*'s eye was a constant 164 cd/m². A circular paper spot of known diameter was attached to *S*'s lower eyelid as a calibration for pupil measurement from photographic enlargements.

Results and Discussion

Pupil diameter during heating and recovery is shown in Figure 4, where each data point represents the mean pupil diameter of the series of three photographs.

According to Hakerem (1967), spontaneous changes in pupillary diameter under dim illumination are of the order of 1–1.5 mm., and typically greater fluctuation oc-

curs with normal light conditions. It is apparent from Figure 4 that there were only the normal spontaneous fluctuations, but no systematic change in pupil diameter with hyperthermia.

According to the Ferry–Porter law, with constant luminance, an increase in pupillary diameter of approximately 58%, more than doubling the retinal illuminance, would be required in order to increase CFF to the degree recorded in Experiment II, i.e., 4 Hz. Because no such change in pupillary diameter was observed as a consequence of hyperthermia in Experiment III, it may be concluded that the increase in CFF was attributable to factors other than a change in pupillary diameter.

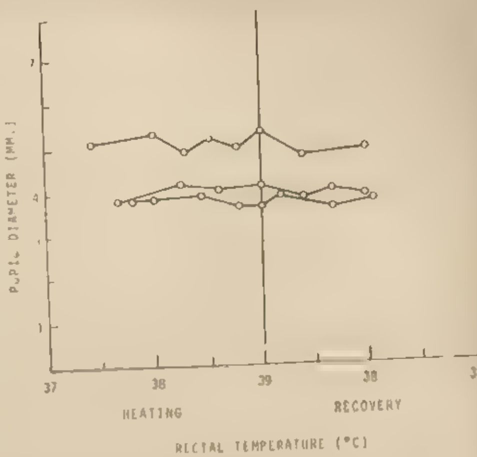


FIGURE 4. Pupil diameter during hyperthermia and recovery for three *Ss*.

General Discussion

The results from both Experiments I and II indicate that temporal acuity increased during hyperthermia. Perception of succession and CFF were both facilitated by increasing body temperature.

In Experiment I, under control conditions, *Ss*' 75% correct response level corresponded to an interflash interval of about 17.3 msec. This value will therefore be considered the average threshold of succession in the present study.

The average threshold of succession may be used to estimate the duration of the psychological moment. An interval separating events by half of the duration of the moment should lead to a rate of correct responding that is 50% greater than the rate expected by chance alone. If the first and second events fell within different halves of the same moment, they would be perceived as successive. Since the first event has an equal likelihood of falling any time within the moment, that particular pair of events would be perceived as simultaneous half of the time and as successive half of the time.

If *S* responds at the 75% correct level on a forced-choice basis, his threshold of succession is actually producing correct responses at a rate of 50% greater than the rate (50%) expected by chance alone. From this, it might be reasoned that the threshold of succession is half the duration of the moment. However, this reasoning ignores the fact that the flashes used to demark the interflash interval had themselves a finite duration—10 msec. in this experiment. Neurophysiological evidence indicates that elements in the retina and cortex respond vigorously to both the onset and offset of a light stimulus (see the review by Grossman, 1967, pp. 213–260). Presumably, the first neural events would arise from the onset of the first flash; the final neural events follow the offset of the second flash. Thus, the duration of the second flash, 10 msec., should be added to the threshold of succession, 17.3 msec., as an estimate of half of the duration of the moment. This results in a value of 54.6 msec. as our best estimate of the *S*'s average normal duration of the psychological moment in this experiment.

Our estimate of the psychological moment is in remarkable correspondence with Kristofferson's (1966, p. 7) average estimate of 54.3 msec. The correspondence is even more surprising in that Kristofferson's test of the perception of succession employed long-lasting (2 sec.) and mixed sensory (auditory and visual) pairs of test stimuli.

During hyperthermia, *Ss*' threshold of succession dropped to an average of 15.7 msec., a reduction of 1.6 msec. from control conditions. It was estimated, in the manner described above, that their average moment was shortened to 51.4 msec., a reduction of about 6% from control conditions.

In Experiment II, the CFF was shown to increase as a function of increasing body temperature. Lockhart (1971) reported similar changes in *Ss*' CFFs as they were tested in a hot (49.2° C.) environment, which raised their mean body temperature by .6°C. The failure to demonstrate any systematic effect of hyperthermia on pupillary diameter in Experiment III leads us to interpret these findings as further evidence that temporal acuity varies as a function of body temperature.

The EEG alpha frequency also increased with body temperature in Experiment II, which supports the findings of Hoagland (1936). The relative changes in alpha frequency and CFF may be similar enough (a maximum 7% increase in alpha; a corresponding 6% increase in CFF) to warrant further consideration of a relationship between these variables. However, neither we nor Lockhart (1971) found any significant relationship between changes in alpha frequency and CFF.

It is of interest to note that the mean alpha frequency recorded prior to heating in Condition H was greater than the control values. Though this result was only marginally significant ($p < .10$), it may indicate that anticipation of heating accelerated *S*s' alpha frequency.

In the present experiments, temporal acuity, as judged from tests of the perception of succession and CFF, increased with core temperature. Previous studies, cited above, have shown that temperature exerts a similar influence upon the perception of time passing. Both phenomena may be explained as resulting from a shortening of the psychological moment; and both may reflect the direct action of temperature on a neural pacemaker.

REFERENCES

Baddeley, A. D. Time estimation at reduced body temperature. *American Journal of Psychology*, 1966, **79**, 475–479.

Bell, C. R. Time estimation and increases in body temperature. *Journal of Experimental Psychology*, 1965, **70**, 232–234.

Bell, C. R., & Provins, K. A. Relation between physiological responses to environmental heat and time judgments. *Journal of Experimental Psychology*, 1963, **66**, 572–579.

Benzinger, T. H., & Taylor, G. W. Cranial measurements of internal temperature in man. In J. D. Hardy (Ed.), *Temperature, its measurement and control in science and industry.* Vol. 3, Pt. 3. *Biology and medicine.* New York: Reinhold, 1963.

Brown, J. L. Flicker and intermittent stimulation. In C. H. Graham (Ed.), *Vision and visual perception.* New York: Wiley, 1965.

Craik, K. J. W. Theory of the human operator in control systems. II. Man as an element in a control system. *British Journal of Psychology*, 1948, **38**, 142–148.

Fox, R. H., Bradbury, P. A., Hampton, I. F. G., & Legg, C. F. Time judgment and body temperature. *Journal of Experimental Psychology*, 1967, **75**, 88–96.

Fraisse, P. *The psychology of time.* New York: Harper & Row, 1963.

Grossman, S. P. *A textbook of physiologi[...] psychology.* New York: Wiley, 1967.

Hakerem, G. Pupillography. In P. H. V[...]bles & I. Martin (Eds.), *A manual of psycho[...]siological methods.* New York: Wiley, 1967

Harter, M. R. Excitability cycles and c[...]ical scanning: A review of central intermitte[...] in perception. *Psychological Bulletin*, 196[...] **68**, 47–58.

Hoagland, H. The physiological control of [...]dgments of duration: Evidence for a chemical [...]k. *Journal of General Psychology*, 1933, **9**, 26[...]87.

Hoagland, H. Pacemakers of human brain w[...]es in normals and in general paretics. *Ame[...]an Journal of Physiology*, 1936, **116**, 604–624.

Kleber, R. J., Lhamon, W. T., & Goldstone, S. Hyperthermia, hyperthyroidism, and time ju[...]ment. *Journal of Comparative and Physiolog[...]al Psychology*, 1963, **56**, 362–365.

Kristofferson, A. B. A time constant involv[...]d in attention and neural information processi[...] (Tech. Rep. CR-427) Washington, D. C.: N[...]tional Aeronautics and Space Administrati[...] 1966.

Kristofferson, A. B. Attention and psyc[...] physical time. In A. F. Sanders (Ed.), *Attenti[...] and performance.* Amsterdam: North-Holla[...] 1967.

Legg, C. F. Alpha rhythm and time judgment[...] *Journal of Experimental Psychology*, 1968, 7[...] 46–49.

Lockhart, J. M. Ambient temperature and th[...] flicker-fusion threshold. *Journal of Experiment[...] Psychology*, 1971, **87**, 314–319.

McReynolds, P. Thinking conceptualized in term[...] of interacting moments. *Psychological Review*, 1953, **60**, 319–330.

O'Hanlon, J. F., Danish, J. E., & McGrath, J. J. Body temperature and the rate of subjective tim[...] (Tech. Rep. 719-8). Goleta, Calif.: Human Factors Research, 1968.

Shallice, T. The detection of change and the perceptual moment hypothesis. *British Journal of Statistical Psychology*, 1964, **17**, 11–35.

Stroud, J. M. The fine structure of psychological time. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955.

Treisman, M. Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychological Monographs*, 1963, **77**(13, Whole No. 576).

von Békésy, G. V. A new audiometer. *Acta Oto-Laryngologica*, 1947, **35**, 411–422.

Walter, W. G. The twenty-fourth Maudsley lecture: The functions of the electrical rhythms in the brain. *Journal of Mental Science*, 1950, **96**, 1–31.

White, C. T. Temporal numerosity and the psychological unit of duration. *Psychological Monographs*, 1963, **77**(12, Whole No. 575).

Wiener, N. *Cybernetics.* New York: Wiley, 1948.

(Received July 24, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 795–798

JUDGMENT OF RECENCY FOR PICTURES AND WORDS[1]

GARY L. LASSEN, TERRY C. DANIEL, AND NEIL R. BARTLETT[2]

University of Arizona

An experiment was conducted comparing judgment of recency (JOR) for line drawings and printed words representing common objects. In a mixed design, groups of 15 *Ss* were tested using either pictures or words at either of two presentation rates. Upon an item's second occurrence, which was marked to be easily distinguished as a test, *Ss* were required to make a judgment of recency, their best estimate of the actual lag for that item. It was found that as lag length increased, accuracy of recency judgment declined. More importantly, an interaction was found between stimulus type and rate of presentation. Words were found to be superior to pictures at a rate of 2-sec. exposure per stimulus, while this pattern was reversed at the 4-sec. exposure rate. The results are discussed in reference to a memory-strength interpretation of JOR and the dual-coding theory of Paivio.

How is the judged recency of a remembered event determined? Recent accounts have related judged recency of an event to the memory strength for that event. Memory strength is usually assumed to be maximal at the time of occurrence of the event and then to decline over time. The stronger the memory trace for an event, the more recently it will be judged to have occurred. The most explicit memory-strength interpretation was presented by Hinrichs (1970; Hinrichs & Buschke, 1968), who extended the basic strength and decision processes proposed in models of event memory to the judgment of recency (JOR). He proposed a two-process memory-strength theory of JOR that described the data for recency judgments of letters reasonably well.

A characteristic finding in both recall and recognition memory paradigms is that information presented pictorially is less susceptible to forgetting than is information presented verbally (Paivio, 1971; Shepard, 1967). This has been interpreted as indicating slower decline (decay) of memory strength for pictorial information than verbal information. Pictorial and verbal materials are known to produce different effects in a variety of other paradigms as well.

The dual-coding theory of Paivio (1971) distinguishes between sequential and parallel accessing of information in memory; the hypothetical verbal system being suited to sequential processing, while the image system is a parallel processing system that organizes input spatially. The theory generally predicts superiority of words to pictures in serial tasks, but negligible differences between pictures and words are expected if the pictures could have been easily labeled.

Pictures have been found to be superior to words in serial tasks using recency paradigms. Yntema and Trask (1963), for example, found discrimination of relative recency (DR) of two unrelated stimuli to be less accurate when the items were nouns than when they were pictures. Using the same paradigm, Fozard (1970) presented pictures and words in the same sequence at a rate of 1.5 sec. per item. He found the best recency discrimination when both items compared were pictures. Further, Fozard and Weinert (1972) found superior JOR performance for pictures when a self-paced procedure was used. Even though *Ss* may have had time to verbally code the picture stimuli in these experiments, the superiority of pictures to words in these tasks presents some difficulty for Paivio's (1971) conceptualizations.

---

[1] This research is based in part on a thesis submitted in partial fulfillment of the requirements for the master's degree in the Department of Psychology at the University of Arizona.

[2] Requests for reprints should be addressed to Neil R. Bartlett, Department of Psychology, University of Arizona, Tucson, Arizona 85721.

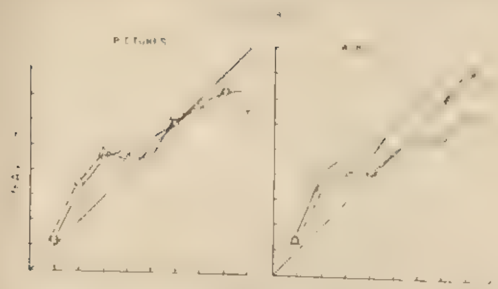


FIGURE 1. Mean judgment of recency (JOR) as a function of Rate of Presentation × Lag Length treatment combination for pictures and words separately.

The present experiment investigated judgment of recency for pictures and words. Picture or word stimuli were presented serially at either a fast (2.0 sec. per stimulus exposure) or a slow rate (4.0 sec. per stimulus exposure). On signaled test trials *Ss* judged the recency of occurrence of the indicated item. The rate factor is relevant to memory-strength interpretations of JOR (e.g., Hinrichs, 1970). The slower rate of presentation results in longer intervals between items and should result in generally longer judgments of recency than the faster rate of presentation. Further, if memory strength for pictorially presented items decays less rapidly over time than does memory for verbal items, the slower rate of presentation should have less effect on JOR for picture items than for words. Also, at either rate of presentation, picture stimuli may be judged to have occurred more recently than word stimuli at the same lag. With regard to Paivio's (1971) hypothesis, rate of presentation would not be expected to produce marked effects as the fastest, 2.0-sec. exposure, rate is sufficiently slow that all *Ss* should be able to code verbally the picture stimuli.

METHOD

Design and subjects. Two between-*Ss* variables, type of stimulus (picture or word) and stimulus exposure time (2 sec. or 4 sec.) were combined with two within-*Ss* variables, trials and lag length (number of items plus one intervening between two occurrences of the same stimulus) into a 2 × 2 × (5 × 9) factorial design. The sixty *Ss* (30 males and 30 females) were volunteers from introduct[illegible] psychology courses at the University of [illegible]zona. Groups of *Ss* were randomly assigned to [illegible]her a "picture" or a "word" stimulus conditio[illegible] [illegible]d to either a 2-sec. or 4-sec. stimulus exposure c[illegible]tion.

Stimuli. Line drawings of familiar objec[illegible] [illegible]sted for prompt common naming, and their corr[illegible]nding word names were photographed and mo[illegible] [illegible] on 35-mm. slides. Stimuli were projected onto [illegible] [illegible]ar-projection screen by a modified Kodak [illegible] [illegible]usel (Model 801) slide projector. Random lists [illegible] [illegible]ach class of stimuli were formed with the restricti[illegible] [illegible]hat no item was reused until at least 10 differen[illegible] [illegible]ms intervened (see Hinrichs, 1970). Five subli[illegible] [illegible]ere formed according to the stimulus ordering te[illegible] [illegible] of Nickerson and Brown (1963). This or[illegible] [illegible]ng technique insured exactly one test at each of [illegible]gs 1 through 9 within each sublist. These sublists [illegible]ere presented to *S* one after another, without pau[illegible] as one continuous list.

Procedure. Each *S* was instructed to mak[illegible] [illegible]n estimate of the lag between two occurrences of [illegible] same stimulus item at the time of its second o[illegible] rence. An item's second occurrence was cle[illegible] designated as a test. The correct recency judgm[illegible] they were instructed, consisted of the numbe[illegible] intervening items since the most recent non[illegible] presentation of that item including itself, or t[illegible] number of intervening items plus one. The *S*s [illegible] sponded verbally, giving "judgments of recency" upon each occurrence of a test. A short practi[illegible] session using similar stimuli was always given [illegible] make sure *S* understood the requirements of th[illegible] task.

RESULTS

Mean JORs at each lag are presented in Figure 1 for all four experimental conditions. In spite of the apparent lack of rate of presentation effect for picture stimuli, there was a general tendency for *Ss* to underestimate the true lag at the 2.0-sec. rate and to overestimate the lag at the 4.0-sec. rate. The main effect of rate was statistically reliable, $F(1, 56) = 4.97$, $p < .05$. Inspection of Figure 1 strongly suggests that rate of presentation had different effects for pictures and words, but the Rate × Stimulus Type interaction was only marginal, just failing to reach the .05 level of significance. Apparently the lack of rate differences for either pictures or words at the short lags (one to four) prevented the rather dramatic differences at the longer lags from emerging in the analysis of variance. This interpretation gains support from the significant Rate × Lag Length interaction, $F(8, 448) = 2.19$, $p < .05$,

indicating that the principal locus of rate effects was at the longer lags.

There was a small but consistent tendency for judgments to become larger as a function of trials (experience), F (4, 224) $= 6.27, p < .01$. Judgments became larger on later trials, but the differences involved were small and followed no significant trend. A significant Trials × Lag Length effect was found, F (32, 1972) $= 3.12, p < .01$. Again, no consistent trend was apparent in the data.

Accuracy of JORs was assessed in terms of absolute deviation of JOR from actual lag. These data can be seen in Figure 1 by comparing mean JOR with true lag indicated by the solid line diagonal. Mean absolute deviations for picture and word stimuli at each presentation rate are also presented in Table 1.

Analysis of variance revealed a significant Stimulus Type × Rate of Presentation effect for mean absolute deviations from actual lag, F (1, 56) $= 5.22, p < .05$. The most notable difference was between words at the slow and fast rates. Pictures differed considerably less as a function of rate. Subsequent post hoc testing (Tukey HSD—Kirk, 1968) disclosed that the fast-word condition was significantly more accurate than the fast-picture condition. The slow-picture condition was significantly more accurate than the slow-word condition. The fast-word and the slow-picture conditions were not significantly different.

TABLE 1

MEANS (AND STANDARD DEVIATIONS) OF ABSOLUTE DEVIATIONS FROM ACTUAL LAG

Stimulus type	Rate of presentation	
	2 sec.	4 sec.
Pictures	1.812 (1.349)	1.479 (1.555)
Words	1.516 (1.316)	1.976 (1.748)

DISCUSSION

The observed superiority of words over pictures in accuracy of JOR at the faster rate of presentation appears to be consistent with Paivio's dual-coding theory. Words arouse the verbal code that is crucial to performance on sequential tasks more readily than do pictures. This difference is usually restricted to cases where the rate of presentation is quite fast, reducing the possibility that pictures would be verbally coded. Contrary to this analysis, pictures were found to be superior to words in JOR accuracy at the slow rate of presentation. This finding is difficult to explain if verbal coding is assumed to mediate performance on sequential memory tasks. While the slow rate of presentation surely provided ample time for *S*s to verbally label the pictures, it seems unlikely that the verbal code for the pictures was more effective in the JOR task than the verbal code for the words themselves. The availability of a verbal code does not appear sufficient to explain all of the observed differences in JOR accuracy for pictures and words. Perhaps the verbal coding system is preferred when rate demands are high, but "imaginal" coding is more effective at slower rates of presentation in spite of the large serial processing component of the JOR task.

The picture–word differences also present difficulty for Hinrichs' (1970) memory-strength interpretation of JOR. The slower presentation rate did not lead to longer JORs for pictures as was predicted by the memory-strength interpretation. The slower rate actually led to more accurate JORs than the faster rate. It may be that contrary to Hinrichs' memory-strength hypothesis, high memory strength for an event need not produce a judgment that the event occurred recently. Instead, high memory strength may produce a more accurate judgment of when the item occurred. The *S*s presented with pictures in the JOR task apparently made beneficial use of the extra 2-sec. stimulus exposure provided by the slow rate of presentation to overcome the effects of longer delays. The *S*s presented with words apparently did not benefit sufficiently from the extra processing time to overcome the detrimental effects of longer delays.

An extension of Morton's (1969) "logogen" model of memory offered by Seymour (1973) may prove useful in explaining picture–word differences as a function of rate. The model proposes that visual inputs for words and pictures are distinguished early in processing and undergo different coding processes. If the ultimate response is vocal, the route for

words is more direct than that for pictures Word input, after having been graphically analyzed, proceeds directly to an exit node where it is available for vocal output. To reach this same node, picture input must first be pictorially analyzed, proceed to a pictorial exit node, and then be interpreted in semantic memory before reaching the vocal exit node. Because there are additional stages in the encoding of pictures, words would have an advantage over pictures in vocalization tasks particularly when fast rates are demanded. These same additional stages may contribute to superior performance with pictures in other memory tasks.

REFERENCES

Fozard, J. L. Apparent recency of unrelated pictures and nouns presented in the same sequence. *Journal of Experimental Psychology*, 1970, **86**, 137–143.

Fozard, J. L., & Weinert, J. R. Absolute judgments of recency for pictures and nouns after various numbers of intervening items. *Journal of Experimental Psychology*, 1972, **95**, 472–474.

Hinrichs, J. V. A two-process memory-str[...] theory for judgments of recency. *Psychol[...] Review*, 1970, **77**, 223–233.

Hinrichs, J. V., & Buschke, H. Judgme[...] recency under steady-state conditions. *Jo[...] of Experimental Psychology*, 1968, **78**, 574–57[...]

Kirk, R. E. *Experimental design: Procedu[...] the behavioral sciences.* Belmont, Calif.: Br[...] Cole, 1968.

Morton, J. Interaction of information in wor[...]ognition. *Psychological Review*, 1969, **76**, 165[...]

Nickerson, R. S., & Brown, C. R. A sti[...] ordering technique for controlled lag recog[...] memory experiments. *Psychological Reports* [...] **13**, 319–322.

Paivio, A. *Imagery and verbal processes.* [...] York: Holt, Rinehart and Winston, 1971.

Seymour, P. H. K. A model for reading, nam[...] and comparison. *British Journal of Psychol[...]* 1973, **64**, 35–49.

Shepard, R. N. Recognition memory for w[...] sentences, and pictures. *Journal of Verbal L[...]ing and Verbal Behavior*, 1967, **6**, 156–163.

Yntema, D. B., & Trask, F. P. Recall as a s[...] process. *Journal of Verbal Learning and [...] Behavior*, 1963, **2**, 65–74.

(Received August 1, 1973)

urnal of Experimental Psychology
)74, Vol. 102, No. 5, 799-805

PROACTIVE INTERFERENCE AND DIRECTED FORGETTING IN SHORT-TERM MOTOR MEMORY[1]

LESLIE BURWITZ[2]

Department of Physical Education and Recreation, Acadia University, Wolfville, Nova Scotia, Canada

Motor proactive interference (PI), produced by requiring *S*s to recall five linear lever displacements in reverse order after a 90-sec. retention interval, was completely eliminated by an instruction to forget the first four positions given prior to the presentation of the criterion (fifth) response and significantly reduced by a similar instruction given immediately before recall. The ability of *S*s to intentionally forget prior motor responses was assumed to explain the inconsistent findings of short-term motor memory research, although it appeared that to-be-forgotten prior items were not completely "discarded," since they were recalled equally as well as to-be-remembered prior items. The discussion centered on the cognitive control of motor behavior and, in particular, on the mechanisms of set differentiation, selective rehearsal, and selective search.

The importance of eliminating irrelevant information was known to the ancients (Cicero, 1959, p. 427). More recently Bjork (1972, p. 218) has suggested that intentional forgetting is fundamental to the efficient functioning of man, since without it he would "degenerate to a proactive-interference-induced state of total confusion." Intentional or directed forgetting has received considerable experimental attention recently in the area of verbal short-term memory (for reviews see Bjork, 1972; Epstein, 1972) but it has not been considered within the realm of motor behavior. This experimentation has led to an increased understanding of verbal memory and one might expect that motor memory research would also benefit from a systematic study of the effect of instructions to forget. The present study was the first such investigation and, since it was specifically designed to test the effect of instructions to forget prior motor learning, the results were relevant to the understanding of short-term motor memory (STMM) proactive interference (PI).

The initial research concerned with the effect of prior motor learning on the retention of a motor response was conducted by Adams and Dijkstra (1966). A Peterson–Perterson (Peterson & Peterson, 1959) paradigm was employed whereby *S* was required to learn and recall seven different linear displacements in turn and, in contrast to the results of verbal short-term memory (STM) research (Keppel & Underwood, 1962) there was no evidence of PI over trials. The null PI finding stimulated further research and, although Montague and Hillix (1968) and Schmidt and Ascoli (1970) reached essentially the same conclusion, Ascoli and Schmidt (1969), Stelmach (1969), and Williams (1971) reported conflicting evidence, in that criterion (final) response recall was indirectly related to the number of prior motor learning movements. The initial motivation for the present research was a desire to resolve this inconsistency, which appeared to be a function of procedural differences. That is, the studies that found no evidence of PI used a Peterson–Peterson learn–recall paradigm (single-trial design), whereas those that were successful in their efforts to produce PI required *S* to learn and

[1] This paper was submitted as a PhD dissertation to the Physical Education Department at the University of Illinois. The author wishes to acknowledge with thanks the insightful comments and suggestions of committee chairman Jack A. Adams and committee members Michael J. Ellis, Rainer Martens, and William E. Montague.

[2] Requests for reprints should be sent to Leslie Burwitz, Department of Physical Education and Recreation, Acadia University, Wolfville, Nova Scotia, Canada.

recall in reverse order a number of different movements within one trial.

The reverse order or multiple-trial design was introduced by Ascoli and Schmidt (1969) to permit the inclusion of a number of motor items into STM very rapidly. Ascoli and Schmidt reasoned, on the basis of a verbal STM study conducted by Bjork, LaBerge, and Legrand (1968), that, because verbal PI is reduced by an instruction to forget previous items, prior motor learning may not interfere with a criterion response if *S* is allowed to "discard" previous traces from STM. They suggested that motor PI is not produced in a single-trial design because *S* is allowed to intentionally forget prior motor learning. To some, this argument may seem irrational. After all, there is ample evidence that directed forgetting affects verbal learning and yet PI is produced in single-trial design experiments in verbal learning (Keppel & Underwood, 1962). The difference between the findings of motor and verbal single-trial design experiments indicates, however, that there is an inherent danger in considering that one set of principles governs behavior in both domains. The single-trial verbal STM studies generally test the retention of trigrams (three items) on a dichotomous (right or wrong) scale, while the single-trial motor STM experiments use only single items and measure retention on a continuous (graded in millimeters or $\frac{1}{8}$-in. units) scale. Thus, the verbal and motor single-trial experiments are not identical and an explanation that does not account for the data in one domain may very possibly explain behavior in the other area.

The ability of *S* to intentionally forget prior motor learning is unknown but, if it is possible, one would appear to be justified in concluding that the inconsistent STMM PI findings are a function of procedural differences which permit or restrict intentional forgetting of prior learning. That is, if *S* can make use of a forget instruction to eliminate motor PI, one would not expect to find PI in a single-trial design, where *S* is allowed to learn, recall, and completely forget each response prior to the pres ta- tion of the next distance, whereas one w ld expect to find PI in a multiple-trial de n, because *S* is required to remember all is- placements in reverse order and is give no opportunity to forget prior learning.

The underlying assumption of the pre nt research was that motor PI, which oc rs in a multiple-trial design, is a manif a- tion of cognitive processes, which o ur consequent with or subsequent to encod . To test this hypothesis the mechanis s, which have been proposed to expl in directed forgetting in verbal STM, w considered to determine which, if any, w relevant to the study of motor mem The six mechanisms, which were subje to analysis, have been referred to erasure, set differentiation, selective hearsal (Bjork et al., 1968), recall lo d (Reed, 1970) or anticipatory recycli g (Epstein, 1969), initial registration (Ree , 1970), and selective search (Epstein, 197 Support for any of the hypotheses w assumed to lend credence to the view th motor behavior is controlled by cogniti processes.

Method

Experimental Design

Thirty *S*s were randomly assigned to each of th 10 independent treatments of a 2 × 5 factorial de sign. The two levels of the first factor involved retention intervals of 5 and 90 sec. The five levels of the second factor were referred to as PI, forget late (FL), forget early (FE_1 and FE_5), and control (CO). The PI treatment involved four prior learning (PL) responses, a criterion response, and an instruction to remember all five responses in reverse order at recall. The FL group received four PL, a criterion response, and an instruction to forget the first four responses presented immediately prior to recall. The two FE treatment groups were given four PL, a criterion response, and an instruction to forget the first four responses given prior to the presentation of the criterion. The FE_1 treatment group was reminded immediately prior to recall to reproduce only the criterion response, whereas in the FE_5 condition *S*s were instructed to recall all five responses in reverse order at that time. Learning and subsequent recall of only one response acted as CO. Absolute and algebraic error scores were computed for the criterion response and for the prior learning (first four) responses recalled in the PI and FE_5 conditions.

Apparatus

The task employed involved linear lever-positioning. The *Ss* were able to hear movements of the almost frictionless lever but they were restricted from viewing the apparatus by means of opaque goggles. The apparatus consisted of two 72-cm. long Thomson case-hardened steel rods, which were mounted at their ends into aluminum blocks so that the rods were horizontal, parallel, and 5-cm. apart. The aluminum blocks were secured onto a 78 × 22 × 1 cm. aluminum base plate, which was in turn screwed to a wooden table of normal height. A slide consisting of four Thomson ball-bushing sleeves (each 3-cm long), an aluminum plate, and a steel handle (8-cm. high) moved freely along the trackway with a range of 0-47 cm. Eight holes, located in the baseboard, accepted a steel pin, which effectively arrested carriage movements at 5, 10, 15, 20, 25, 30, 35, and 40 cm. The criterion distance was 20 cm. and a scale marked in millimeters was used by *E* to measure error indicated by a fine pointer.

Experimental Procedure

On entering the experimental room, *S* was seated in front of the apparatus so that his right shoulder was opposite and a full arm's length from the right-hand edge of the carriage. A demonstration (47 cm.) and a standard set of instructions were given. The instructions informed *S* that he would either have to make one or five movements down to a stop and then, after a rest, he would have to reproduce some or all of the movements without the assistance of the stop. The *Ss* were told that they would not be given instructions as to which movements they would have to remember until later in the experiment and they were informed that the best procedure was to try to remember every position until they heard otherwise.

On the command "grasp" *S* took hold of the handle and waited for the instruction "move." On this signal *S* moved the handle at a moderate, deliberate pace from right to left until he hit the stop. He maintained his grasp on the handle for a period of 3 sec. at which time he was instructed to "relax," which meant he was to place his hand on a pad located on the table in front of him. The sequence of events described above defined a learning trial and for *Ss* in the CO treatment the retention interval followed. After the retention interval *S* was given recall instructions. The commands "regrasp" and "estimate" acted as signals for *S* to move the lever to where he thought it had been during learning.

If *S* received PI, he made four movements before the sequence described above. The prior positions were 5 or 10 cm. longer or shorter than the criterion distance of 20 cm. The number of possible permutations of previous positions was 24. The order of previous positions was randomized but each of the 24 permutations occurred approximately as often. The intertrial interval was 10 sec. The stop, defining the distance of each movement, was positioned by *E*, and he was also responsible for all return movements of the handle to the start position.

The forget instruction, given in the FE_1, FE_5, and FL treatments, informed *S* that he was to completely erase the first four responses and concentrate only on the last (criterion) response. The FL instruction was given in a 5-sec. period immediately prior to recall, and FE instructions were presented in the last 5 sec. of the 10-sec. intertrial interval between the fourth and the criterion response. The recall instructions were presented in a 5-sec. period immediately prior to recall. The PI and FE_5 *Ss* were told to recall all five responses in reverse order (10-sec. intertrial interval), FL and FE_1 *Ss* were instructed to recall only the criterion response, and CO *Ss*, to reproduce their one position. Following the final recall trial, a postexperimental interview was conducted to determine what strategy *S* used, what he thought he did with recall and forget items, and what effect the forget instruction had had.

Subjects

The *Ss* were 300 (153 male, 147 female) right-handed volunteers (average age, 19 yr., 9 mo.) from the general student body at Acadia University, Nova Scotia. There were approximately an equal number of males and females in each treatment, the largest discrepancy being an overload of 16 to 14 in either direction.

RESULTS

Absolute Error

Criterion response. The analysis of absolute error indicated that *Ss* were on average 2.93 cm. off target. There was no difference between the performance of males (3.00) and females (2.88). The Retention Interval × Treatments (2 × 5) analysis revealed that the retention interval and treatment main effects were significant, $F(1, 290) = 37.05$, $p < .001$, and $F(4, 290) = 8.97$, $p < .001$, respectively. The *Ss* recalled more accurately at 5 sec. (2.21) than after a 90-sec. retention interval (3.57), and a Newman Keuls test indicated that both the FL (3.32) and PI (3.99) treatments recalled significantly (FL at .05 level; PI at .01 level) less than CO (2.23), FE_1 (2.37), and FE_5 (2.55) conditions. None of the other differences was significant.

The Retention Interval × Treatment interaction was also significant, $F(4, 290) = 3.34$, $p = .01$. Figure 1 graphically illus-

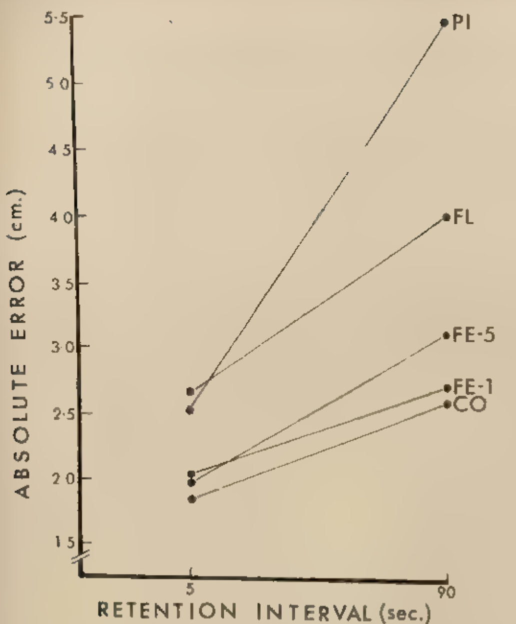


FIGURE 1. Criterion response: Retention Interval × Treatment interaction. (Abbreviations: PI = proactive interference, FL = forget late, FE_1 = forget early, remember one, FE_5 = forget early, remember five, and CO = control.)

trates the interaction and a post hoc Newman–Keuls test indicated that significant recall decrements occurred 5–90 sec. under PI ($p < .01$) and FL ($p < .05$) but not under FE_5, FE_1, and CO. There were no differences between the five treatments at 5 sec., but at 90 sec. PI recall (5.44) was significantly ($p < .01$) worse than FL (4.00), FE_5 (3.12), FE_1 (2.69), and CO (2.60) recall. In addition, FL recall was significantly ($p < .05$) worse than that of the FE_1 and CO treatments.

Prior learning responses. Four treatments (PI at 5 and 90 sec. and FE_5 at 5 and 90 sec.) contributed data to the Treatment × Retention Interval × Trials (PI, FE_5 × 5, 90 sec. × 1–4) repeated measures analysis. The treatment and retention interval main effects were not significant, F (1, 116) = 2.79, $p > .05$, and F (1, 116) <1, respectively. There was no difference between PL recall of *Ss* who were trying to remember the first four responses (PI = 5.10) and those who had been instructed to forget prior positions (FE_5 = 6.28).

The main effect for trials was significant, conservative F[3] (1, 116) = 4.76, p .05. A post hoc Newman–Keuls test indicated that Trial 1 (4.34) was significantly different from Trial 2 (5.80, .05 level), Trial 3 (6.48, .01 level), and Trial 4 (6.1[illegible] .05 level).

The Treatment × Retention Interval × Trials analysis indicated that none of the interactions was important, with the exception of Retention Interval × Trials, which was marginally significant, conservative F (1, 116) = 3.02, $p < .10$. A post hoc Newman–Keuls test revealed that none of the eight cells differed significantly from each other (all ps > .05).

Algebraic Error

Criterion response. The analysis of algebraic error data indicated that *Ss* tended to undershoot the criterion very slightly (−.11 cm.). There was no difference between the performance of males (.03) and females (−.25). The Retention Interval × Treatment (2 × 5) analysis revealed no significant effects (all Fs < 1).

Prior learning responses. The Treatment × Retention Interval × Trials (2 × 2 × 4) repeated measures analysis revealed a slight tendency to overshoot (1.9 cm.) on Trials 1–4. The treatment, retention interval, and trials main effects were not significant, F (1, 116) < 1, F (1, 116) = 1.62, $p > .10$, and conservative F (1, 116) = 1.57, $p > .10$, respectively, and neither were any of the interactions.

Discussion

The directed forgetting paradigm employed in the present study contributed in a general and positive way to the understanding of motor memory. The results support the hypothesis that the inconsistency in STMM PI research is a function of procedural differences, which either permit or restrict intentional forgetting of prior learning. Proactive interference was produced by using a multiple-trial design, where *S* was not allowed to forget the first four responses since he was required to

[3] The conservative F test was used for within-*Ss* data to guard against probable positive bias of standard F ratios.

reproduce them after the criterion response at recall. The effect of PL was completely eliminated by an early instruction to forget the first four positions (FE) as well as being substantially reduced by a similar instruction not presented until immediately prior to recall (FL). Thus, Ascoli and Schmidt (1969), Stelmach (1969), and Williams (1971) were able to produce PI by adopting a multiple-trial design, which restricts intentional forgetting of prior learning, while Adams and Dijkstra (1966), Montague and Hillix (1968), and Schmidt and Ascoli (1970) found no evidence of PI because they used a single-trial design that contained an implicit cue to forget previous learning prior to the presentation of each criterion response.

The PI reduction following forget instructions indicates that motor PI is not a consequence of prior motor activity per se. Motor PI was not affected by the mere act of responding to or encoding prior movements but by operations, which can be assumed to be cognitively based, since they occurred contiguous with and/or subsequent to encoding. Verbal researchers have referred to these operations as erasure, set differentiation, selective rehearsal (Bjork et al., 1968), recall load (Reed, 1970) or anticipatory recycling (Epstein, 1969), initial registration (Reed, 1970), and selective search (Epstein, 1972). The present study was designed to determine which, if any, of the above mechanisms was relevant to the study of motor behavior and, although further research is necessary to substantiate the findings, the interaction data indicated that the mechanisms of selective rehearsal, set differentiation, and selective search were the most influential.

Analysis of the interaction revealed that the FE_1, FE_5, and FL treatments reduced PI, that there were no differences between any of the five treatments at 5 sec., and that there was a significant difference between FE_1 and FL conditions. The largest reduction in PI occurred under FE_1. Each of the various hypotheses is able to account for the reduction. The forget early cue may have allowed *S* to (*a*) completely "discard" prior learning (erasure), (*b*) tag or organize recall items in such a way as to separate them from forget items (set differentiation), (*c*) concentrate rehearsal time on the criterion response (selective rehearsal), (*d*) recall the criterion without having to recycle prior learning for subsequent recall (anticipatory recycling), (*e*) properly encode the criterion at the time of its presentation (initial registration), and/or (*f*) reduce the size of the search set from which the criterion was to be recalled (selective search).

The second largest reduction in PI occurred under FE_5. The finding discriminates between the six hypotheses offered to explain directed forgetting. The erasure, set differentiation, selective rehearsal, and initial registration notions are able to account for the difference but the result contradicts the predictions of the anticipatory recycling and selective search hypotheses. The anticipatory recycling hypothesis postulates that having to recycle other material for subsequent recall is detrimental to criterion response recall. Thus, it predicts no difference between FE_5 and PI, because both conditions required *S* to recycle four prior learning responses during criterion response recall. One might argue that the first four responses were stronger under PI because FE_5 *Ss* were instructed to forget prior learning and, although this argument seems plausible, it is undermined by the fact that the first four positions were recalled with equal facility under the two conditions. Apparently, the to-be-forgotten items did not interfere at recall in the manner outlined by the anticipatory recycling hypothesis despite being strong enough to do so. Additional evidence against the recycling notion was provided by the nonsignificant difference between FE_1 and FE_5. Evidently the size of the recall load did not affect directed forgetting.

The nonsignificant difference between the first four responses under FE_5 and PI provides strong evidence against the erasure hypothesis and support for the set differentiation and selective search notions. The to-be-forgotten positions were not completely "discarded" but rather they appeared to be stored separately (set differentiation) and recalled equally as well as the to-be-remembered items, when *S* was aware within which set to search (selective search).

The set differentiation and selective search mechanisms are considered by Epstein (1972) as necessary conditions for directed forgetting in verbal learning. The present data indicate that there is no reason to doubt that partitioning the input is fundamental to directed forgetting in motor learning but the selective search mechanism does not appear to be as vital within the realm of motor behavior. In verbal learning, Epstein, Massaro, and Wilder (1972) and Shebiske and Epstein (1972) report that forget instructions are only effective when there is a difference in search-set

size. In the present study, search-set size was equated using a procedure much the same as that employed by Shebilske and Epstein, and the difference between the FE_5 and PI treatments indicates that forget instructions reduce motor PI in the absence of search-set differences. The finding contradicts the search-set hypothesis and suggests that the selective search mechanism may be a necessary condition for directed forgetting in verbal learning but not motor learning.

Of the four hypotheses capable of explaining the reduction of PI under FE_5, the erasure notion appears to be the weakest, since forget items were not completely "discarded." The initial registration, selective rehearsal, and set differentiation explanations are considered below. The initial registration hypothesis postulates that prior motor learning impairs criterion response acquisition and that this, rather than interference, causes the PI effect. The hypothesis is able to account for the reduction of PI under the two forget early conditions but it cannot explain the fact that none of the treatments differed significantly from each other at 5-sec. recall. If one assumes that immediate recall (5-sec.) reflects original learning, the equality of the conditions at this point in time indicates that prior learning does not influence criterion response acquisition. In addition, the enhanced encoding position is unable to explain the reduction of PI following the FL cue. The difference between PI and FL treatments contradicts the idea that PI reduction is a consequence of an event that precedes the presentation of the criterion response.

The difference between the PI and FL treatments also proves to be difficult for the selective rehearsal notion to explain. The hypothesis postulates that PI is a consequence of divided rehearsal time. Having to share rehearsal attention over the retention interval is assumed to reduce the probability of the correct response being accurately reproduced. The FL and PI treatments had essentially the same rehearsal opportunities and consequently one would expect there to be no difference between them. A proponent of the rehearsal position might argue that the difference between PI and FL treatments could be attributed to the extra few seconds of rehearsal time FL *S*s had between the FL cue and recall. This possibility, although it cannot be dismissed, seems unlikely as a few seconds would hardly appear to be long enough to produce such a difference.

The significant reduction of PI under FL can be explained by the erasure, anticipatory recycling, selective search, and set differentiation hypotheses. The first three explanations have been discussed above and weaknesses have been pointed out in each. The set differentiation hypothesis is also unable to explain all the data. The significant difference between FE_1 and FL conditions contradicts Bjork's (1970) suggestion that the ability to tag or organize recall and forget items is unaffected by the placement of the forget cue. The finding does not necessarily contradict the idea that partitioning of the input must be assumed in any account of directed forgetting (Epstein, 1972), but it does indicate that FE_1 *S*s either originally learn the criterion response better than FL *S*s (initial registration) or that they are better able to retain the criterion response because they are allowed to concentrate their rehearsal time on that response rather than having to divide it between five different positions (selective rehearsal). The difference in initial registration was not apparent at 5 sec. and, although any difference in original learning may interact with the length of the retention interval, it would appear that the selective rehearsal hypothesis is better able to explain the difference between FE_1 and FL. Further research is obviously necessary to substantiate this point.

It is clear from the above discussion that, when *S* is instructed to forget a number of prior motor responses, he responds by invoking more than one of the mechanisms described above. The set differentiation, selective search and selective rehearsal operations were considered the most important, as, in combination, they were able to explain all of the data without the assistance of the erasure, anticipatory recycling, and initial registration hypotheses. In addition, the latter three hypotheses received no direct support as forget items were not completely forgotten, an increased recall load did not prove to be detrimental to recall, and there were no differences between the five treatments at 5 sec.

In conclusion, the present study indicates that the inconsistent findings in STMM PI research are a function of procedural differences which either permit or restrict intentional forgetting of prior learning. That is, motor PI is produced not merely by the occurrence of prior learning but by requiring *S* to remember five movements and reproduce them all in reverse order after a 90-sec. retention

nterval. It appears that, when S is able to orget prior learning, he voluntarily sets aside prior, potentially interfering responses, selectively rehearses the criterion response, and finally, selectively searches either the recall or forget set to find the appropriate response, which is subsequently accurately reproduced regardless of which store it is located in. Thus, the mechanisms that operate in motor memory are assumed to be cognitively oriented and, although this may seem dubious to those who view the mental and motor domains as completely separate, there are others (Adams, 1971) who appreciate that higher order processes are involved in even the simplest motor behaviors.

REFERENCES

Adams, J. A. A closed-loop theory of motor learning. *Journal of Motor Behavior*, 1971, **3**, 111–150.

Adams, J. A., & Dijkstra, S. Short-term memory for motor responses. *Journal of Experimental Psychology*, 1966, **71**, 314–318.

Ascoli, K. M., & Schmidt, R. A. Proactive interference in short-term motor retention. *Journal of Motor Behavior*, 1969, **1**, 29–36.

Bjork, R. A. Positive forgetting: The non-interference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 255–268.

Bjork, R. A. Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. New York: Wiley, 1972.

Bjork, R. A., LaBerge, D., & Legrand, R. The modification of short-term memory through instructions to forget. *Psychonomic Science*, 1968, **10**, 55–56.

Cicero. *De Oratore*. Book II, chap. LXXIV, lines 298–300. (Trans. by E. W. Sutton & H. Rackham) (Rev. ed.) Cambridge: Harvard University Press, 1959.

Epstein, W. Poststimulus output specification and differential retrieval from short-term memory. *Journal of Experimental Psychology*, 1969, **82**, 168–174.

Epstein, W. Mechanisms of directed forgetting. In G. H. Bower (Ed.), *The psychology of learning and motivation*. Vol. 6. New York: Academic Press, 1972.

Epstein, W., Massaro, D. W., & Wilder, L. Selective search in directed forgetting. *Journal of Experimental Psychology*, 1972, **94**, 18–24.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Montague, W. E., & Hillix, W. A. Intertrial interval and proactive interference in short-term motor memory. *Canadian Journal of Psychology*, 1968, **22**, 73–78.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

Reed, H. Studies of the interference process in short-term memory. *Journal of Experimental Psychology*, 1970, **84**, 452–457.

Schmidt, R. A., & Ascoli, K. M. Intertrial intervals and motor short-term memory. *Research Quarterly of the American Association for Health, Physical Education and Recreation*, 1970, **41**, 432–438.

Shebilske, W., & Epstein, W. Effect of forget instructions with and without the conditions for selective search. Cited by W. Epstein, Mechanisms of directed forgetting. In G. H. Bower (Ed.), *The psychology of learning and motivation*. Vol. 6. New York, Academic Press, 1972. Pp. 171–175.

Stelmach, G. E. Prior positioning responses as a factor in short-term retention of a simple motor task. *Journal of Experimental Psychology*, 1969, **81**, 523–526.

Williams, I. D. The effects of practice and prior learning on motor memory. *Journal of Motor Behavior*, 1971, **3**, 205–212.

(Received August 3, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 806–812

REGRESSION EFFECT AND INDIVIDUAL POWER FUNCTIONS OVER SESSIONS[1]

ROBERT G. WANSCHURA AND WILLIAM E. DAWSON[2]

University of Notre Dame

Stevens attributed individual differences in power function exponents mainly to differences in "the regression effect" and suggested that balancing for regression may largely do away with individual differences. Cross-modality matches of loudness to apparent time duration, and vice versa, were repeated over six sessions for each of 20 *S*s with 24 hr. between sessions. Individual differences, indicated by analysis of variance, Kendall's *W*, and significant positive correlations across sessions, occurred before and after regression balances. Group exponents for duration matches increased over sessions, but those for loudness matches and those obtained by regression balance did not change significantly.

The development of direct scaling techniques, e.g., magnitude estimation and cross-modality matching, has shown that sensory magnitude grows as a power function of stimulus intensity when data are averaged for a group of *S*s (Stevens, 1961, 1971). In addition, there is evidence that the power law may provide an adequate description of the data of individual *S*s (Marks & Stevens, 1966; Stevens & Guirao, 1964), although some investigators (Luce & Mo, 1965; Pradhan & Hoffman, 1963) find the description a poor one. Even when individual data yield power functions, exponents vary widely across *S*s. These individual differences appear to be reliable effects, since when a group of *S*s repeats a task like magnitude estimation, *S*s' later exponents are correlated with their earlier ones (Rule, 1966).

Luce (1972), noting the wide variation in exponents, has doubted the possibility of finding invariant relations among measures in psychophysics as well as the development of a "coherent system of units as has been done in physics [p. 100]." He concludes that individual differences do not allow the specification of a unit of sensory magnitude apart from the individual.

Various explanations for idiosyncratic exponents have been offered. Rule (1966) and Rule and Markley (1971) interpreted them as reflecting consistent individual response strategies. They feel, as do others (Jones & Marcus, 1961), that *S*s tend to use a characteristic range of the variable under their control with some *S*s using a narrow range and others more extended ones. Ekman, Hosman, Lindman, Ljungberg, and Akesson (1968) pointed to differences in perceptual sensitivity as well as to response biases when not all of their variance due to *S*s was accounted for by the latter.

Teghtsoonian and Teghtsoonian (1971) have suggested that session-to-session correlations are transitory and are mainly due "to chance factors when the constraints introduced by repeated judging are minimized [p. 149]." They find that such correlations are not significant for intersession intervals greater than 24 hr. and argue that a sensory or judgmental trait of individual *S*s would not disappear in this manner.

Stevens (1971) suggested that individual differences are due largely to the regression effect (Stevens & Greenbaum, 1966). This effect is a tendency for the person to use a shortened range of the variable under his control, whether it be the number continuum as in magnitude estimation, or a physical continuum as in cross-modality matching. The shortening takes place about the average value of the variable

---

[1] Based on a master's thesis completed by the first author at the University of Notre Dame. Also, presented as a paper at the meeting of the Midwestern Psychological Association, Chicago, May 1973.

[2] Requests for reprints should be sent to William E. Dawson, Department of Psychology, University of Notre Dame, Notre Dame, Indiana 46556.

under control. Such a tendency to "regress" toward the mean value of a series of controlled stimuli was first noted by Hollingworth (1910).

Because of this effect, when one continuum is matched to another, two different power functions are obtained depending on which is the adjusted variable. Each of these functions is a straight line in a log–log plot but one has a steeper slope than the other. To present both functions on the same plot, the ordinate and the abscissa are interchanged for that curve obtained by adjusting the variable represented on the abscissa. This necessitates drawing a line with a slope equal to the reciprocal of the exponent of the regularly fitted power function. Indow and Stevens (1966) have argued that the most representative function that describes the relation between the two continua is a power function with an exponent that is the geometric mean of the exponents for the two power functions in such a plot. Such an exponent is considered balanced for the regression effect and is obtained directly as the geometric mean of the exponent for the ordinate continuum and the reciprocal of the exponent for the abscissa continuum.

No one to date has thoroughly investigated whether balancing for regression will do away with reliable differences in exponents. The present research attempts to answer this question, and furthermore it examines whether or not these effects are stable with repetition of the scaling task over several sessions. Stevens (1971) reported that on some occasions exponents may change with repetition. Teghtsoonian and Teghtsoonian (1971) found that exponents for apparent length and apparent area were steady over five sessions and Mitchell and Gregson (1971) reported individual differences in olfactory exponents that were stable over four sessions. Otherwise, not much has been reported about repetition effects for more than two sessions.

Two questions were of primary focus: (*a*) Do consistent individual differences occur across sessions? If so, do these differences disappear when the exponents are balanced for regression in judgments? (*b*) Are group exponents stable over repeated sessions or do they change with repetition? If they change, is this change symmetrical for both tasks, i.e., does it disappear when the regression effect is balanced out using the geometric mean?

Apparent time duration and loudness were chosen as modalities primarily for the reproducible regression effects that have been obtained with them previously. Cross-modality matches were undertaken in which loudness was adjusted to match given time intervals and, in turn, apparent time duration was adjusted to match given levels of loudness.

Method

Subjects. The *S*s were 20 naive volunteers, 16 of whom were undergraduate or graduate students attending summer school at the University of Notre Dame. Three *S*s were wives of graduate students and 1 was an undergraduate not attending summer school. Ten men and 10 women participated.

Apparatus. For matches of loudness to given time intervals, an audio oscillator produced a 1,000-Hz. tone whose output level was controlled by a variable attenuator and a sone potentiometer (Poulton & Stevens, 1955) in series with a pair of earphones. Each *S* listened to the tone while adjusting the potentiometer for the appropriate loudness. The phones were calibrated so that the voltage measured across them could be converted to sound pressure readings in dynes/cm². Eleven intervals were used: .30, .42, .60, .84, 1.20, 1.70, 2.40, 3.40, 4.80, 6.80, and 9.60 sec. Time durations were controlled by a Hunter timer, which lighted a small white lamp for the appropriate interval.

When apparent duration was the adjusted variable, *S* turned on the lamp for the desired duration by pressing a key, and the duration was recorded by a Standard timer. The levels of loudness matched were produced by 11 discrete sound pressure levels (re .0002 dynes/cm²), which varied 40–90 db. in 5-db. steps.

Procedure. Each *S* participated in six sessions scheduled at the same time each day over 6 consecutive days. In a session, *S* matched both duration to loudness and loudness to duration with half of the *S*s first making all the time matches and then all of the loudness ones. The remaining half began by adjusting loudness and then continued with duration. The order in which each modality served as the matched variable was randomly assigned to both *S*s and sessions under the constraints that, for any *S*, half of the sessions began with duration and, for any session, half of the *S*s began with duration. Within any session, the stimuli were presented in random order with *S* making only one match to each

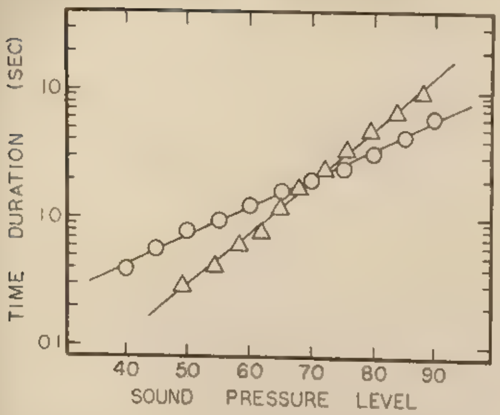


FIGURE 1. Cross-modality matches of apparent duration to loudness (circles), and vice versa (triangles), for the data pooled over sessions and *S*s.

stimulus level. For each part of each session an appropriate set of instructions was read by *S* before he began.

RESULTS

Geometric means of the matches made to each stimulus were computed for the data pooled over (*a*) *S*s and sessions, (*b*) *S*s for each session, and (*c*) sessions for each *S*. Power functions were then fitted to each of these sets of geometric means for both duration and loudness adjustments. Time and sound pressure, respectively, were the independent variables in the equations fitted. In addition, power functions were fitted to each *S*'s matches for each session.

Group data. Figure 1 shows that the group data combined over sessions is described well by power functions for both cross-modality matches. The duration exponent (DE), for the matches where time was adjusted, is indicated by a slope of .45. The loudness exponent (LE) was 1.24, and hence the figure gives t e reciprocal of this value, .81, for the slo where loudness is the adjusted variable. sence of a regression effect is indicate v the difference in slope for the two tches. The regression angle (RA) between e two fitted lines amounts to .259 radian Taking the geometric mean of .45 and . gives .60 as the balanced exponent (BE

Previous studies place the expo t for apparent duration at 1.10 (Stevens 1961) or, more recently, at 1.03 (Stevens & reenbaum, 1966). For loudness, exp ients from .60 (Stevens, 1961) to .68 (St vens, 1969) have been reported. Thus, th predicted slope of the equisensation fu tion lies in the interval from .60/1.10, or . 5, to .68/1.03, or .66. If values of .64 for oudness and 1.03 for duration are taken the predicted slope for the cross-mo lity match is .62, a value only slightly l rger than the .60 obtained as the BE.

Analyses of the data for each session are summarized in Table 1. The DEs show an orderly increase over sessions, but the LEs are more stable. Again, most of the BEs lie in the interval .55–.66. Figure 2 gives plots of the geometric means of the matches for each session, and these results are well described by power functions as indicated by the straight lines fitted to the datum points.

Individual data. Random samples of the results of individual *S*s are plotted in Figures 3 and 4. The results, combined over sessions, for six *S*s chosen at random are given in Figure 3; those for six *S*s in single sessions are given in Figure 4. The individuals vary considerably in their manner of responding in the cross-modality matches. Data combined over sessions

TABLE 1

SESSION MEANS FOR THE FOUR MEASURES

Measure	Session						
	1	2	3	4	5	6	*M*
Duration exponents	.38	.44	.43	.45	.49	.51	.45
Loudness exponents	1.21	1.33	1.20	1.22	1.24	1.22	1.24
Balanced exponents	.59	.60	.61	.62	.66	.67	.62
Regression angles (in radians)	.36	.27	.32	.28	.26	.25	.29

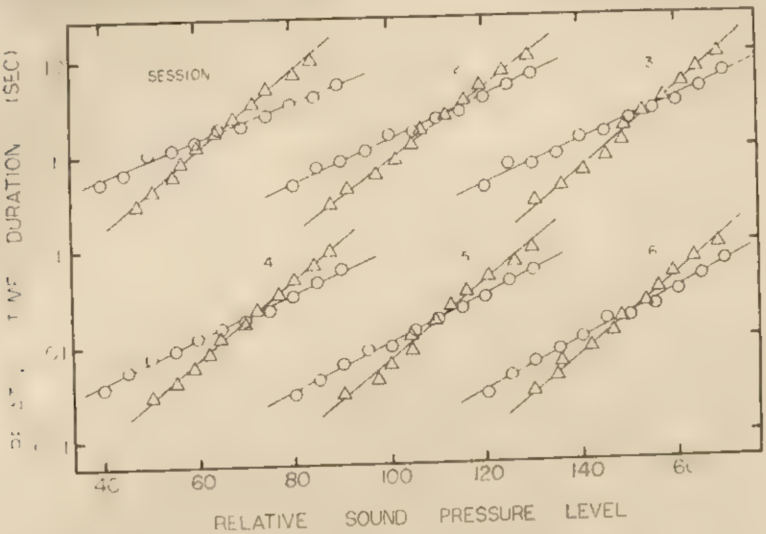


FIGURE 2. Cross-modality matches of apparent duration to loudness (circles), and vice versa (triangles), for the six sessions. (All pairs of curves, except the upper left, have been shifted to the right or downward for presentation.)

tend to approximate a power function better than data for one session.

Balanced exponents for each *S* in each session were obtained by taking the geometric mean of the appropriate duration and inverse loudness exponents (ILEs). In addition, for the same 120 combinations of session and *S*, RAs were computed.

Changes over sessions. To assess the stability of the various measures, a Treatments × *Ss* analysis of variance was applied to each matrix of 120 measures. Thus, four analyses in all were conducted, one each for DEs, LEs, BEs, and RAs. The sessions effect was significant when duration was adjusted, F (5, 95) = 5.56, $p <$

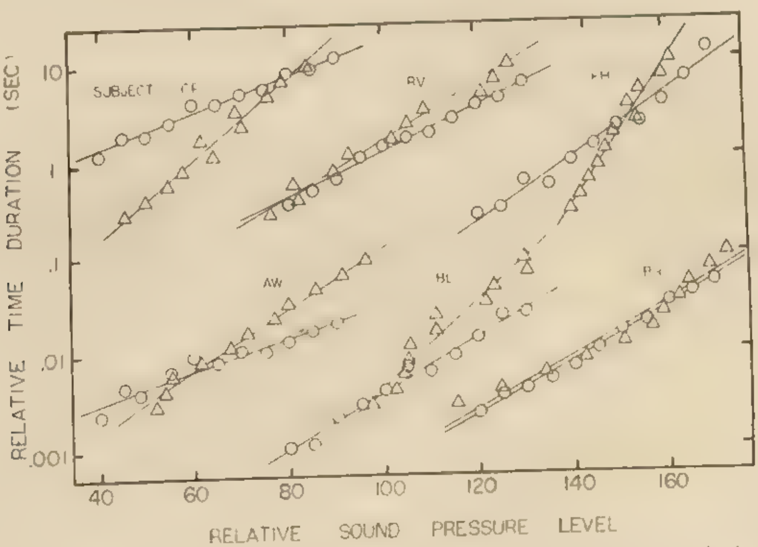


FIGURE 3. Matches of apparent duration to loudness (circles), and vice versa (triangles), for each of six randomly selected *Ss* with their data pooled over sessions. (All pairs of curves, except the upper left, have been shifted to the right or downward for presentation.)

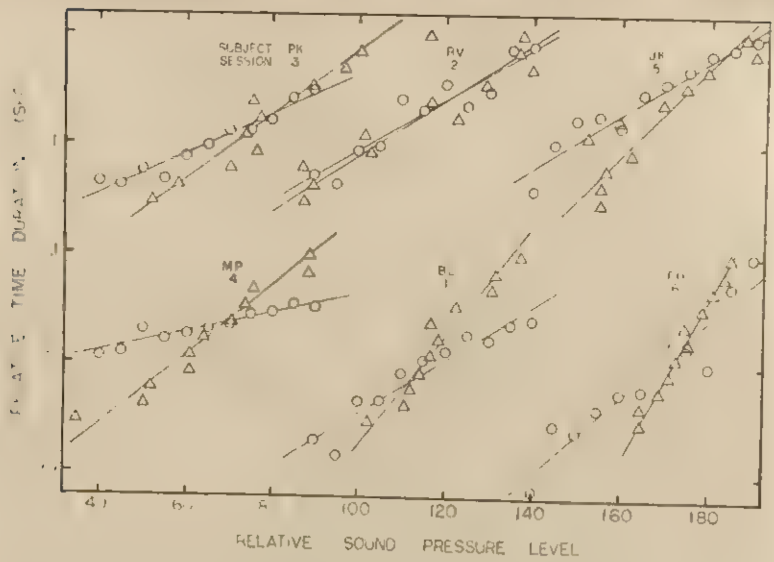


FIGURE 4. Single session matches of apparent duration to loudness (circles), and vice versa (triangles), for each of six randomly selected *Ss*. (All pairs of curves, except the upper left, have been shifted to the right or downward for presentation.)

.001, but not when loudness was, F (5, 95) = .862, $p > .05$. The BEs for the group data tended to increase over sessions but not significantly, F (5, 95) = 2.14, $.05 < p < .1$. On the other hand, RAs changed reliably, F (5, 95) = 3.62, $p < .01$.

Additional comparisons were undertaken for those measures where a sessions effect occurred. Duncan's new multiple range test indicated that the DE for Session 1 was significantly less than those for Sessions 5 and 6, but other comparisons of session means yielded no difference. With RA means, Session 1 was greater than Session 6, but other comparisons were not significantly different.

Individual differences. Tukey's test for nonadditivity (see Myers, 1972) was carried out for each of the four measures. Nonadditivity effects were not significant with F (1, 94) = 2.98, 1.54, 2.47, and <1 for DEs, LEs, BEs, and RAs, respectively. Thus, it was permissible to test the *Ss* effect for each of the measures. In all cases this effect was significant at the .001 level: DEs, F (19, 95) = 9.73; LEs, F (19, 95) = 13.18; BEs, F (19, 95) = 12.87; and RAs, F (19, 95) = 10.29.

The fact that the Tukey tests found the additive model to be appropriate implies that *Ss* were not differentially affected by sessions. In order to examine the consistency of *S* differences over sessions, Kendall's coefficient of concordance was calculated for each of the four measures. All were significant at the .001 level with W (19) = .64, .74, .72, and .67 for DEs, LEs, BEs, and RAs, respectively. Similar evidence for persisting individual differences was given by intersession product-moment correlations (df = 18), all of which were positive. For DEs, all 15 possible rs were significant at the .05 level using a one-tailed test and all but 2 were so at the .025 level. With LEs, all rs were significant at the .01 level, while for BEs, all were reliable at the .05 level with all but 2 being at the .005 level. The rs for RA were all significant at the .05 level, and all but 2 were at the .025 level. Table 2 presents the intersession correlations for successive pairs of sessions. The Session 1 vs. Session 2 values tend to be lower than succeeding correlations.

Correlational analysis was also used to examine further Stevens's contention that

individual exponents may be mainly due to regression. Under Stevens's hypothesis, a negative correlation is expected between individual duration exponents and their corresponding inverse loudness exponents. That is, if regression balances for all individuals are to lead to the same exponent, then an *S* with a low DE must have a compensatingly high ILE and vice versa. Cross-modal Pearson *r*s were calculated between these individual exponents both for each session and using individual exponents based on all sessions. The correlations between modalities at each session were all positive, with only Session 2 significant, r (18) = .51, $p < .05$. The correlation based on combined sessions was not significant, r (18) = .30, $p > .1$.

Discussion

In general, results provide additional support for the use of direct scaling methods and the assertion of the power law for both the data of groups and individual *Ss*. The presence of regression in judgments was apparent, and balances for this effect yielded several interesting results.

Stability of the measures. A change in exponents with repetition occurred for apparent duration but not for loudness. The DE change was in the direction of decreased regression over sessions, and it occurred gradually. Since the balanced exponents are based on the DEs and LEs, it is not surprising that they tend to increase over sessions even if not significantly so. Although the BEs are, in effect, stable, the RAs indicate that regression decreases significantly over sessions. Stevens and Greenbaum (1966) suggested that regression balances may not be perfectly stable when they stated that

> the distorting constraints that affect one variable, say, magnitude estimation, may be different from the constraints that affect the other variable, which may involve, for example, the turning of a knob that has peculiar characteristics of its own. It seems unlikely that the errors in two such different tasks as magnitude estimation and magnitude production would inevitably cancel exactly and completely [p. 446].

A possible reason for the difference between the sessions effects for DEs and LEs lies in the manner in which the two continua were adjusted. Loudness adjustment was continuously variable in both directions of intensity, but duration adjustment was only possible in an increasing direction. Loudness matches did not require return to a "zero" intensity before increasing or decreasing the intensity. With duration, *Ss* had to start over in order to produce a longer or shorter interval. There are two ways in which such adjustment differences could be avoided. First, *Ss* could be required to make loudness matches in a manner analogous to duration ones, that is, by increasing monotonically from zero loudness for all settings. Another approach would be to use continua which are usually adjusted in a similar fashion, e.g., time duration and force of handgrip or loudness and brightness. Whether some types of adjustment lead to stable exponents, while others do not, deserves to be investigated.

TABLE 2

Intersession Correlations for Successive Pairs of Sessions

Measure	Adjacent sessions				
	1–2	2–3	3–4	4–5	5–6
Duration exponents	.45	.56	.66	.77	.68
Loudness exponents	.65	.84	.72	.78	.86
Balanced exponents	.71	.78	.74	.67	.85
Regression angles	.39	.69	.54	.66	.69

Individual differences. The results, in agreement with earlier research, demonstrate that consistent individual differences persist over sessions. Since only one match per stimulus for each continuum was made per session and since there were approximately 24 hr. between sessions, it was expected that the intersession correlations, especially for Session 1 vs. Session 2, would be low and possibly nonsignificant. Teghtsoonian and Teghtsoonian (1971) found nonsignificant correlations in such a situation for magnitude estimation data. The present results, however, were obtained with cross-modality matching which conceivably could make for the difference. For both continua, all entries in the correlation matrices and the *Ss* effects of their analyses of variance were significant, indicating that there is a consistency in responding even with a 24-hr. intersession interval. The highly significant Kendall's *W*'s indicated reliability of individual differences over repetitions.

The attempt to remove individual differences by balancing for regression was not successful. They were still present with BEs as indicated by the analysis of variance, the *W*,

and the intersession correlations. Other sources of idiosyncratic exponents must still exist—perhaps several causes will be found responsible. The fact that the average intersession correlation for BEs was about the same as those for the DEs and LEs suggests that the proportion of variance attributable to individuals is similar in all three cases.

The positive cross-modal correlations between DEs and ILEs argue against Stevens's expectation that individual differences will largely vanish after a regression balance. As mentioned earlier, in order for all individual BEs to converge on the same value, a DE that is relatively low must be accompanied by a relatively high ILE, and so forth. Thus, negative correlations would be expected. Incidentally, they would also be expected by a "general range hypothesis." According to such a hypothesis, idiosyncratic exponents would be due to consistent use by individuals of a relatively constant proportion (large or small) of the range of any adjusted variable as they go from continuum to continuum. Some investigators (Rule, 1966; Rule & Markley, 1971) have suggested that the individual differences arise from the idiosyncratic use of the dependent variable—a sort of response style. According to a general range hypothesis, such response styles would be expected to transfer across continua.

What other possible sources of *S* differences might there be? Stevens (1961) earlier suggested three components of variability: *S*'s modulus, *S*'s conception of a subjective ratio, and the operating characteristic of *S*'s sense organ. Stevens (1971) rules out the latter because normal sense organs do not vary as widely as individual exponents. Differences in the modulus can be ruled out since they presumably affect the multiplicative constant but not the exponent of a power function. Differences in subjective ratios clearly would affect the exponent and it is known that they occur. Svenson and Akesson (1966), for example, find such differences but they also show that *S*s differ in other ways as well. Accounting for individual variability will probably entail discovering at least several factors and assessing how they operate in combination as Stevens (1971) has suggested.

REFERENCES

Ekman, G., Hosman, B., Lindman, R., Ljungberg, L., & Akesson, C. Interindividual differences in scaling performance. *Perceptual and Motor Skills*, 1968, **26**, 815–823.

Hollingworth, H. L. The central tendency of judgment. *Journal of Philosophy, Psychology and Scientific Methods*, 1910, **7**, 461–469.

Indow, T., & Stevens, S. S. Scaling of saturation and hue. *Perception & Psychophysics*, 1966, **1**, 253–271.

Jones, F. N., & Marcus, M. J. The subject effect in judgments of subjective magnitude. *Journal of Experimental Psychology*, 1961, **61**, 40–44.

Luce, R. D. What sort of measurement is psychophysical measurement? *American Psychologist*, 1972, **27**, 96–106.

Luce, R. D., & Mo, S. S. Magnitude estimation of heaviness and loudness by individual subjects: A test of a probabilistic response theory. *British Journal of Mathematical and Statistical Psychology*, 1965, **18**, 159–174.

Marks, L. E., & Stevens, J. C. Individual brightness functions. *Perception & Psychophysics*, 1966, **1**, 17–24.

Mitchell, M. J., & Gregson, R. A. M. Between-subject variation and within-subject consistency of olfactory intensity scaling. *Journal of Experimental Psychology*, 1971, **89**, 314–318.

Myers, J. L. *Fundamentals of experimental design.* Boston: Allyn & Bacon, 1972.

Poulton, E. C., & Stevens, S. S. On the halving and doubling of the loudness of white noise. *Journal of the Acoustical Society of America*, 1955, **27**, 329–331.

Pradhan, P. L., & Hoffman, P. J. Effect of spacing and range of stimuli on magnitude estimation judgments. *Journal of Experimental Psychology*, 1963, **66**, 533–541.

Rule, S. J. Subject differences in exponents of psychophysical power functions. *Perceptual and Motor Skills*, 1966, **23**, 1125–1126.

Rule, S. J., & Markley, R. P. Subject differences in cross-modality matching. *Perception & Psychophysics*, 1971, **9**, 115–117.

Stevens, J. C., & Guirao, M. Individual loudness functions. *Journal of the Acoustical Society of America*, 1964, **36**, 2210–2213.

Stevens, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961.

Stevens, S. S. On predicting exponents for cross-modality matches. *Perception & Psychophysics*, 1969, **6**, 251–256.

Stevens, S. S. Issues in psychophysical measurement. *Psychological Review*, 1971, **78**, 426–450.

Stevens, S. S., & Greenbaum, H. B. Regression effect in psychophysical judgment. *Perception & Psychophysics*, 1966, **1**, 439–446.

Svenson, O., & Akesson, C. A. Fractional and multiple estimates in ratio scaling. *Reports from the Psychological Laboratories, University of Stockholm*, 1966, No. 202.

Teghtsoonian, M., & Teghtsoonian, R. How repeatable are Stevens's power law exponents for individual subjects? *Perception & Psychophysics*, 1971, **10**, 147–149.

(Received August 10, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 813–817

PRESENTATION RATE AND INTRALIST REPETITION EFFECTS IN IMMEDIATE PROBE RECALL[1]

V. DAVID BURNS[2]

University of Maryland

Recall in an immediate memory study was tested by a missing-item probe for two levels of presentation rate (1 and 4 items per second) and four intralist repetition conditions. The presentation rate results contradicted the single-process forgetting hypotheses that were considered, including constant or variable time decay and interference-only. Repetition effects supported a multistore or multiprocess hypothesis. The divergent performance patterns exhibited by groups of Ss who were exposed to different presentation rate orders emphasize the importance of separating strategy from memory effects.

Two important issues which have been investigated in the study of short-term memory are sources of forgetting (time decay versus interference) and intralist repetition effects. These issues have ordinarily been examined in isolation from each other and hence it is reasonable to investigate their interaction; however, both are related to the more general problem of whether multiple memory systems (e.g., Atkinson & Shiffrin, 1968) or multiple encoding or retrieval processes (e.g., Craik & Lockhart, 1972; Murdock, 1972) are necessary to describe human performance in the immediate memory paradigm. This latter question was examined in the present study by testing the predictions of a number of "single process" models either for presentation rate or intralist repetition effects.

With respect to the presentation rate variable, model predictions are based on the following assumptions: (*a*) that forgetting rates as a function of intervening items are most appropriately measured by sensitivity estimates as prescribed by signal detection theory (see Banks, 1970), (*b*) that *S*s adhere to nonrehearsal instructions when given, and (*c*) that only a single response is required per trial, as in Waugh and Norman's (1965) probe procedure. If it is hypothesized that the only cause of forgetting is interference from subsequently presented items within a list (Norman, 1966), forgetting rates should not differ across presentation rates. Support for this hypothesis has been found by Norman (1966) and has been replicated (Anderson & Burns, 1973) using a missing-item probe with visual presentation (four digits per exposure), required voicing of individual items, and nonrehearsal instructions.

Alternatively, rate of forgetting is predicted to be greatest at the slowest presentation rate according to any of the following hypotheses: pure time decay, a combination of interference and time decay (Wickelgren, 1970), or differential degrees of processing across presentation rates (Massaro, 1970). Wickelgren (1970) has reported an experiment in which this prediction was supported for a recognition task. One purpose of the present study was to replicate Anderson and Burns (1973) with visual presentation of individual list items. Estimates of sensitivity and reaction time for individual *S*s were obtained in order to allow a test of the forgetting rate hypotheses.

The second variable examined in the present study was the frequency and spacing of intralist repetitions. For all three

[1] This research is based on a master's thesis submitted to the Graduate School of the University of Maryland and was completed while the author held a Maryland Fellowship. Support was furnished by the Biomedical Sciences Support Center of the University of Maryland to the Center for Language and Cognition, and computer time was provided in full by the Computer Science Center. The author wishes to thank Nancy Anderson for advice and assistance throughout all stages of the project and John Holmgren and David Horton for critical readings of the manuscript.

[2] Requests for reprints should be sent to V. David Burns, Department of Psychology, University of Maryland, College Park, Maryland 20742.

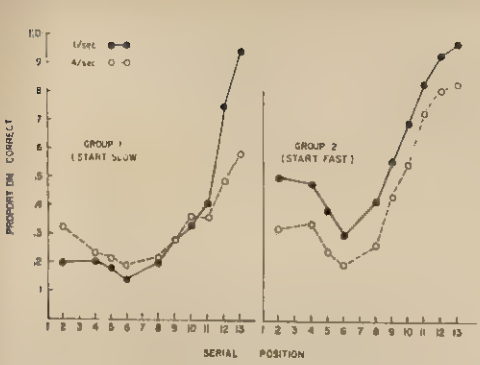


FIGURE 1. Proportion correct as a function of serial position for each rate and group.

previous studies in which no evidence indicating time decay was obtained (Anderson & Burns, 1973; Norman, 1966; Waugh & Norman, 1965), stimulus sequences were used in which the repetition of critical (tested) items was not investigated. It is possible that the failure to find evidence for time decay can be partially attributed to the absence of a control for repetition of critical items. However, repetition effects are important in themselves, since at least three different outcomes are predicted depending on the theoretical model chosen. First, if recall in the immediate memory paradigm reflects the contribution of two distinct memory systems (i.e., recently presented items tend to be recalled from a short-term store whereas earlier items are retrieved from a long-term store), one would expect repetitions to affect performance differentially across serial position. This hypothesis is suggested by the facilitative effect of repetitions found in "long-term" memory studies (e.g., Waugh, 1970) but the absence of any facilitation when a single item is repeated in the "probe memory-span" task (Jahnke, 1969). Second, repetitions may facilitate performance across all serial positions (Bernbach, 1969). Third, repetitions may degrade performance if item-to-item associations mediate recall (e.g., Wickelgren, 1966). According to the associative hypothesis the repetition of an item in a sequence should decrease the probability of recall of the items immediately following the repeated item (if different) because of competing A–B, A–C associations. These three hypotheses were tested by varying the repetition ructure of the stimulus lists.

METHOD

Lists of fifteen digits were composed ·ing all 10 digits (0 to 9) with the restriction t' no item occur more than twice per list. Each l was constructed according to one of the fol ng constraints, which define the four repetitio ditions: (*a*) The critical item was repeated at , of one intervening item (i.e., . . . iji . . . wh i is the critical item and j is any other digit w' is not repeated)—condition CR-1. (*b*) The cr l item was repeated at a lag of six intervenin items—condition CR-6. (*c*) The item in the ser osition preceding the critical item was repeated lag of six intervening items, and the critical ite as not repeated—condition PR. (*d*) Neither t ritical item nor the preceding item was repeat in the list (although other items did repeat)— dition NR. When an item was purposely repeat it was always repeated on only one side of th ritical position. Hence, if j $\geq$ 8 was the critical serial position, the repetition was always placed before the critical item, and conversely for j $<$ 8.

Presentation rates of one and four digits per second (with item "on-times" of approximately 1 and .27 sec., respectively) were chosen to replicate previous studies. On any single trial only one digit was probed and the critical item occurred only in serial positions 2, 4, 5, 6, 8, 9, 10, 11, 12, or 13. Each *S* was exposed to all (80) levels of the three main variables of Presentation rate × Repetition × Serial position. The two levels of presentation rate were varied across sessions, but the 40 possible Repetition × Serial position conditions occurred equally often in a random order within each session. Six University of Maryland undergraduates were recruited as *S*s and paid $2 per hour plus a bonus based on proportion correct. Each *S* participated in nine sessions (120 lists per 50-min. session) at the slow rate and three (320 lists in 75 min.) at the fast presentation rate. As a control for order effects one group of three *S*s (Group 1) was exposed to the slow presentation rate initially, and another group of three *S*s (Group 2) was exposed to the fast presentation rate initially. Because of scheduling conflicts, the only constraint imposed was that Group 1 (start slow) *S*s be exposed predominately to the slow presentation initially, and that Group 2 (start fast) *S*s be exposed predominately to the fast presentation initially. The consequence of this arrangement was that *S*s were exposed to the following sequence of sessions:

Group 1:	Group 2:
SSSSSSSFFSSF (2 *S*s)	FSSFFSSSSSS
SSSSSSSFFSFS (1 *S*)	FSFSFSSSSSSS (1 *S* each)
	FSFFSSSSSSSS.

Stimuli were displayed and responses recorded by means of a CRT terminal connected to the time

sharing computer system at the University of Maryland. After a ready signal on the CRT, *S* initiated the stimulus presentation. Each digit was displayed individually to the right of the preceding digit such that no two items were ever visible simultaneously. Each digit was read aloud as it appeared on the screen, and *Ss* were instructed not to rehearse previous items. The instructions were repeated on numerous occasions throughout the experiment.

Following presentation of the 15 digits, the missing-item probe, consisting of the stimulus list which just occurred except for a single missing digit, was displayed and two hyphens were printed at the position of the item which was to be recalled. One of the 10 digits and a confidence judgment were entered while the probe remained visible, then the screen was cleared and another ready signal given. Two rest intervals lasting a minimum of 1 min. were provided at equal intervals during the session.

Results

Performance as measured by proportion correct is summarized for each presentation rate in Figure 1. An analysis of variance of the arcsin transform of proportion correct indicated that performance at the slow presentation rate was superior, F (1, 4) = 15.8, $p < .05$ (mean difference = .087). Presentation rate interacted with serial position only for Group 1 (start slow), F (9, 18) = 11.5, $p < .001$, which accounts for the significant Groups × Rate × Serial position interaction, F (9, 36) = 3.56, $p < .005$.

Since the forgetting rate predictions for all models are specified in terms of sensitivity estimates, Memory Operating Characteristic (MOC) curves were generated and A_g, a nonparametric estimate of sensitivity representing the area under the MOC, was calculated for each *S* and condition. The sensitivity estimates are plotted in Figure 2 for each group and presentation rate. Performance as estimated by A_g is superior at the slow presentation rate for both groups. (The sharply reduced A_g in serial positions 12 and 13 for Group 2 reflects the inadequacy of this measure when performance is almost perfect.)

The slopes of the linear least-squares fits to log A_g as a function of number of intervening items (the reverse of serial position) were employed as estimates of forgetting rates, as specified by the models

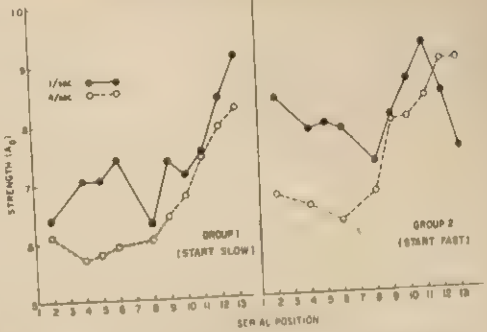


FIGURE 2. Strength (A_g) as a function of serial position for each rate and group.

considered above. (If A_g for serial positions 2 or 4 exceeded the minimum A_g for a given *S* and rate by .1, or if the difference in A_g for positions 12 or 13 and the maximum A_g exceeded .1, those positions were not included in the slope estimate.) The forgetting rates obtained by this procedure were greater at the fast presentation rate for all *Ss*, and the slope differences were significant for four of six *Ss* (two-tailed confidence interval, $p < .05$). The zero-intercepts estimated by least-squares were greater at the fast presentation rate for four of six *Ss*.

Mean reaction times were significantly longer at the slow presentation rate, F (1, 4) = 85.33, $p < .005$, (mean difference = 5.3 sec.). This result has also been obtained by Norman (1966) and Murray (1968).

The overall differences among repetition conditions were not significant but a significant interaction of repetitions with presentation rate and serial position was observed for Group 1, F (27, 54) = 2.83, $p < .001$. Separate analyses performed on the first and last five positions indicated that this interaction was significant only for the first five positions tested, F (9, 19) = 3.52, $p < .025$. Performance was generally superior for condition CR-6 but the differences among conditions varied across presentation rate and position.

For Group 2 (start fast) repetitions interacted with serial position, F (27, 54) = 2.98, $p < .001$, and the differences among repetition conditions were significant only

for the first five positions tested, $F(3, 6) = 6.9$, $p < .025$. For these positions (2, 4, 5, 6, and 8) the critical item was recalled with highest probability when repeated, regardless of lag (conditions CR-1 and CR-6). Recall was poorest for condition NR in which neither the preceding nor the critical item was repeated, and performance was intermediate for lists in which the item preceding the critical item was repeated (condition PR).

Group 2 (start fast) *S*s performed significantly better than Group 1 (start slow) *S*s, $F(1, 4) = 26.86$, $p < .01$ (overall difference in proportion correct = .20), and a similar pattern of group differences was apparent for the A_g measure. The performance superiority exhibited by Group 2 *S*s was restricted primarily to early positions at the slow rate and later positions at the fast rate. Aaronsson (1967) has reported a similar result.

Discussion

The forgetting rate differences obtained in the present study contradict all "single process" models considered, namely, models in which it is assumed that encoding across serial position is constant (for a given presentation rate and nonrehearsal instructions), and that forgetting results either from interference, time decay, or some combination. The forgetting rate differences obtained are consistent with a number of multiprocess models (e.g., Atkinson & Shiffrin, 1968), since performance for recently presented items (later serial positions) is predicted to be independent of presentation rate but recall of items in early positions is assumed to be an increasing function of study time. In addition, the multiprocess model can more easily account for the pronounced primacy effect apparent in Figures 1 and 2 without resorting to post hoc explanations.

Additional evidence that the memory process used by *S*s in this task is not unitary was indicated by the significant effects of repetitions for early but not later serial positions. Although repetitions interacted in a complex manner with presentation rate and serial position for one group, recall was generally facilitated by item recurrence for early but not later serial positions. This supports the hypothesis that the processes of encoding and/or retrieval for early items are analogous to those involved in "long-term" recall studies, whereas retention of later items more closely corresponds to the mnemonic processing which takes place in the memory span task. Neither of the other two hypotheses considered regarding repetition effects were supported.

The pattern of group differences supports the hypothesis suggested by Aaronsson (1967), that *S*s trained at a slow presentation rate tend to adopt an "immediate" perceptual strategy at both presentation rates, whereas *S*s trained at a fast rate tend to adopt a "delayed" identification strategy. However, in the present study it may be more appropriate to hypothesize that the group differences result from a differential delay in "deeper-level" or "long-term memory" processing. Group 1 *S*s (exposed to the slow rate initially) may have attempted to maximize the "depth" of processing of individual items at both presentation rates. However, Group 2 *S*s, trained initially at the fast rate, may have delayed "deeper-level" processing at both presentation rates, since any processing of individual items was impossible at first. The result of utilizing a delayed strategy would be to acquire greater information about the relationship among items, hence Group 2 *S*s would be expected to show superior performance particularly for early serial positions at the slow presentation rate. The "immediate" strategy has the additional disadvantage of requiring more processing time per item, resulting in a greater probability that the encoding of later items at the fast presentation rate is disrupted.

The results of varying both presentation rate and intralist repetition provide further evidence for the necessity of postulating more than a single memory process or dimension to describe retention in the immediate probe recall paradigm. The differences between groups can be accounted for in terms of strategy differences which are compatible with the multiprocess model. The two memory processes hypothesized may be more clearly differentiated by varying the degree of redundancy over more intricately patterned stimulus sequences.

References

Aaronsson, D. Temporal factors in perception and short-term memory. *Psychological Bulletin*, 1967, 67, 130–144.

Anderson, N. S., & Burns, V. D. A comparison of presentation rates using a missing-item probe

test of immediate memory. *Bulletin of the Psychonomic Society*, 1973, **2**, 200–202.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. II. New York: Academic Press, 1968.

Banks, W. P. Signal detection theory and human memory. *Psychological Bulletin*, 1970, **74**, 81–99.

Bernbach, H. A. Replication processes in human learning and memory. In J. T. Spence & G. H. Bower (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. III, New York: Academic Press, 1969.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671–684.

Jahnke, J. G. The Ranschburg effect. *Psychological Review*, 1969, **76**, 592–605.

Massaro, D. W. Perceptual processes and forgetting in memory tasks. *Psychological Review*, 1970, **77**, 557–567.

Murdock, B. B., Jr. Short-term memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. Vol. V. New York: Academic Press, 1972.

Murray, D. J. Repeated recall in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 358–365.

Norman, D. A. Acquisition and retention in short-term memory. *Journal of Experimental Psychology*, 1966, **72**, 369–381.

Waugh, N. C. On the effective duration of a repeated word. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 587–595.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–105.

Wickelgren, W. A. Associative intrusions in short-term recall. *Journal of Experimental Psychology*, 1966, **72**, 853–858.

Wickelgren, W. A. Time, interference and rate of presentation in short-term recognition memory. *Journal of Mathematical Psychology*, 1970, **7**, 219–235.

(Received August 11, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 818-823

EFFECTS OF POSITIVE AND NEGATIVE FORCE-CONTINGENT REINFORCEMENT ON THE FRUSTRATION EFFECT IN HUMANS[1]

GAIL DITKOFF[2] AND RONALD LEY

State University of New York at Albany

Ninety college undergraduates learned a force-discrimination lever-press response under conditions of positive or negative reinforcement. The magnitude of the reinforcement *S*s received for correct training-trial responses was varied. Following training, *S*s received five extinction trials. Consistent with a generalized-drive interpretation of the frustration effect (FE), during early extinction mean response force on a nonreinforced lever (L_2) increased for both the positive-reinforcement and the negative-reinforcement groups. An unexpected finding was a decrease in mean response force on the force-contingent reinforcement level (L_1) from late training trials to early extinction trials by the negative-reinforcement groups. The L_1 and the L_2 data suggested a trend in which the magnitude of the FE increased as the magnitude of reinforcement on training trials increased.

Amsel's (1958) frustration–drive hypothesis states that nonreinforcement following the development of an anticipation of reinforcement results in the aversive motivational state of frustration. According to frustration–drive theorists, frustration acts as a source of drive (D), which results in an increase in the vigor of behavior that immediately follows frustration (the frustration effect—FE). Empirical support for the notion that nonreinforcement following a history of *positive* reinforcement is motivating has been obtained with infrahuman *S*s, using double-runway, hurdle-jump, and operant lever-press apparatuses (Amsel & Rousel, 1952; Daly, 1971; Marzocco, 1951), and with humans, using force, movement speeds, and latency of lever-press responses as dependent variables (Blixt & Ley, 1969; Ryan & Watson, 1968). With respect to *negative* reinforcement, a comparatively small amount of research has been reported. Although Lambert and Hammond (1970) were unable to obtain an FE in rats using escape from shock as a negative reinforcer in a double runway, Graham and Longstreth (1970) did obtain FEs in rats in a double-runway study in which shock escape served as the negative reinforcer. The primary purpose of the present study was to extend the investigation of the effects of termination of negative reinforcement (avoidance of monetary loss) on the FE to the realm of human behavior. In the context of this study, the termination of negative reinforcement during extinction trials consisted of punishment, i.e., monetary loss following responses which had previously been negatively reinforced. On the basis of current theories of the FE, it was hypothesized that the termination of reinforcement following both positively and negatively reinforced training would result in the FE in humans.

The secondary purpose of the present study was to investigate the effect of different magnitudes of both positive and negative reinforcement on the FE. According to both Amsel (1958) and Spence (1960), the magnitude of the FE is related to the strength of *S*'s anticipation of reinforcement (r_G). As the strength of r_G increases, nonreinforcement should produce greater amounts of frustration. Researchers have found that the magnitude of the FE increased (*a*) when the similarity between a runway and its goal box was increased (Amsel & Hancock, 1957), (*b*) when *S*'s nearness to task completion when he was thwarted was increased (Hanner & Brown,

[1] This research was supported in part by a State University of New York Research Foundation Grant in Aid to the second author.

[2] Requests for reprints should be sent to Gail Ditkoff, who is now at the University of Pittsburgh, Johnstown, Pennsylvania 15904.

1955), and (*c*) when the number of reinforced acquisition trials given prior to nonreinforcement was increased (Penny, 1960). According to Spence (1960), another way in which the strength of r_G may be varied is by varying the magnitude of the reinforcement that *S*s receive during training; i.e., *S*s which receive larger reinforcements during training should develop stronger r_G, and hence show greater FEs, than *S*s which receive smaller reinforcements during training. On the basis of this theoretical position, the second hypothesis of the present study was that the magnitude of the FE would be a function of the magnitude of reinforcement (both positive and negative) received during training, i.e., as the amount of reinforcement received for each correct training-trial response increased, the magnitude of the FE would increase.

In the double-runway studies (e.g., Amsel & Rousel, 1952) the use of the same reward (i.e., food) in the goal boxes of both the first runway (R_1) and the second runway (R_2) has led to the question of whether an increase in R_2 response vigor reflects an increase in general drive level or an increase in food-related motivation. Levine and Loesch's (1967) finding that the vigor of a continuously reinforced instrumental response (food-reinforced lever presses) increased during periods when a dissimilar instrumental response (water-reinforced chain pulls) was being extinguished was interpreted as evidence for the hypothesis that nonreinforcement, following continuous reinforcement, leads to increases in D. The tertiary purpose of the present study was to test further the hypothesis that the FE reflects an increase in D, which will be evidenced in an increase in the force of a second response (a response unrelated to the reinforcement contingency of the frustrated response).

By means of a double-lever force-contingent reinforcement apparatus, *S*s successively pressed each of two levers (L_1 and L_2) on each experimental trial. Correct L_1 responses were either positively or negatively reinforced during training trials and were not reinforced during extinction trials. Lever 2 responses were never reinforced. Therefore, a significant L_2 frustration effect (i.e., an increase in the force of L_2 responses following L_1 nonreinforcement) could not be interpreted as the result of the termination of L_2 reinforcement and would support an explanation in terms of the FE reflecting increases in D.

Method

Subjects. The *S*s were 90 undergraduate students enrolled at the State University of New York at Albany.

Apparatus. A modification of the force-contingent reinforcement apparatus designed by Blixt and Ley (1969) was used to measure changes in the force of responses on a force-contingent reinforcement lever (L_1) and a nonreinforced lever (L_2) when a series of continuously reinforced responses was followed by nonreinforcement. These two vertical 48.26-cm.-high steel levers were mounted side-by-side (30.5 cm. apart). There was no discernable movement of the levers when force was applied. The left-hand lever was painted red and the right-hand lever green. Two signal lights (one red and one green) were directly behind and approximately 5 cm. above their correspondingly colored levers. Between the two signal lights, at approximately eye level, was a digit counter (Lafayette Model 57-7-B). At the middle of each lever (front and rear) were two bonded strain gauges (SR-4 Type FA-50-35, 350 ohm resistance, manufactured by Baldwin-Lima-Hamilton) which measured the amount of force applied to the levers. Pressure on the levers resulted in the development of a small signal voltage which was routed through a Carrier amplifier to a Brush eight-channel pen recorder (Model RD-1682-30) and caused a pen deflection on a moving (1 mm/sec) graphical tape. The graphical tape was preexperimentally scaled, and since each lever was connected to a separate pen, a direct reading in grams of force on each lever was possible.

A test control box was manually set at a constant 5-sec. intertrial interval. Beginning with the red light, each of the two signal lights was successively activated for 5-sec. durations. The red-light circuit activated and reset one of two force-band discriminators that monitored voltage output from the strain gauges. Located on Force Band Discriminator 2 was a control switch that enabled *E* to program the L_1 response force-counter activation contingency; i.e., *E* controlled the amount of force that would activate the counter.

Procedure. From the total sample of 90 *S*s, an equal number of *S*s ($n = 15$) was randomly assigned to each of six reinforcement-type (positive reinforcement—PR—or negative reinforcement—NR) by magnitude-of-reinforcement (0¢, 5¢, or 10¢) groups, i.e., PR-0¢, PR-5¢, PR-10¢, NR-0¢, NR-5¢, and NR-10¢. The *S*s were instructed to press L_1 after

TABLE 1

MEAN LEVER 1 AND LEVER 2 RESPONSE FORCE AS A FUNCTION OF TRIALS FOR [illegible] REINFORCEMENT-TYPE BY MAGNITUDE-OF-REINFORCEMENT GROUP

Group	Lever 2				Lever 1			
	Prefrustration trials		Immediate postfrustration trials		Prefrustration trials		Immed. postfrustration trials	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
PR-0¢	5.08	2.18	5.33	2.36	3.72	.23	3.80	.97
PR-5¢	4.63	1.05	4.58	.99	3.92	.41	4.15	.63
PR-10¢	4.35	1.66	4.53	1.92	3.71	.32	4.00	.67
NR-0¢	5.28	1.57	5.32	1.81	3.81	.39	3.18	.61
NR-5¢	4.83	1.95	5.18	2.60	3.45	.42	3.28	1.01
NR-10¢	4.21	.80	4.38	1.29	3.61	.29	3.63	.92

Note. Abbreviations: PR = positive reinforcement, NR = negative reinforcement.

the onset of the L_1 signal light and to press L_2 after each L_1 press, but not until the onset of the L_2 signal light. The *Ss* were further instructed that the function of L_2 was to reset the recording mechanism in L_1.

The PR *Ss* were instructed that following each correct L_1 press (3–5 kg.) a point would accumulate on the counter, while NR *Ss* were instructed that following each incorrect L_1 press (a press less than 3 kg. or greater than 5 kg.) a point would accumulate on the counter. The PR *Ss* were further instructed that each counter point represented 5¢ (Group PR-5¢) or 10¢ (Group PR-10¢) that they had earned, while NR *Ss* were told that they initially had $4.00 to work with and that each counter point represented 5¢ (Group NR-5¢) or 10¢ (Group NR-10¢) which would be deducted from their $4.00. The PR control *Ss* (Group PR-0¢) were told that regardless of their performance they would be paid $1.50 for participating in the study, but that they should try to accumulate points, while NR control *Ss* (Group NR-0¢) were told that they should try not to accumulate points.

Following a practice period, which was designed to familiarize *Ss* with the experimental apparatus and with the proprioceptive feedback associated with a correct L_1 response, training trials were administered until a criterion of five consecutive correct L_1 responses following the tenth training trial was reached. Following training trials five extinction trials were given during which, regardless of the correctness of L_1 responses, no points (additional money) accumulated for the PR groups, whereas one point (a loss of money—punishment) accumulated for the NR groups.

The magnitude of the FE was assessed by comparing the response force of the first L_2 response that followed planned L_1 nonreinforcement (i.e., L_2 force on extinction trial 1—L_2 postfrustration trials) with the mean L_2 response force on the last five training trials (L_2 prefrustration trials), and by comparing the response force of the first L_1 response which followed planned L_1 nonreinforcement (i.e., the L_1 force on Extinction Trial 2—L_1 postfrustration trials), with the mean L_1 force on the last four training trials plus the first extinction trial (L_1 prefrustration trials).

RESULTS

The performance data in terms of mean force of response for each of the six experimental groups (positive or negative reinforcement by 0¢, 5¢, or 10¢ magnitude of reinforcement) on pre- and postfrustration trials for Levers 1 and 2 are given in Table 1. These data were analyzed by means of a four-variable mixed factorial design in which the between-*Ss* variables were reinforcement magnitude (0¢, 5¢, or 10¢) and reinforcement type (positive or negative reinforcement), and the within-*Ss* variables were trials (pre- and postfrustration trials) and levers (L_1—the force-contingent reinforcement lever—and L_2—the reset lever).

With respect to the primary hypothesis, the Reinforcement Type × Trials × Levers interaction was significant, F (1, 84) = 8.70, $p < .005$, thus indicating that the Reinforcement Type × Trials interaction was different from L_1 to L_2. Separate analyses of this interaction revealed that on L_2 the Reinforcement Type × Trials interaction was not significant, F (1, 81) = .14, $p > .05$, whereas on L_1 the Reinforcement Type × Trials interaction was, F (1, 84) = 8.28, $p < .005$. The means of the PR and NR groups on the pre- and postfrustration trials on L_2, which are plotted in Figure 1, clearly indicate an

increase in force of response from pre- to postfrustration trials for both the PR and NR groups, F (1, 84) = 4.11, $p < .05$. That is, with respect to L_2, the FE was obtained under negative as well as positive reinforcement, results consistent with the primary hypothesis. The observed difference between the PR and the NR groups was not significant, F (1, 84) = .10, $p > .05$.

The means of the PR and NR groups on the pre- and postfrustration trials on L_1, which are plotted in Figure 2, indicate an observed increase in force of response from pre- to postfrustration trials for the PR group, findings consistent with the data based on L_2 and with the primary hypothesis, but a decrease in force for the NR group, F (1, 84) = 5.25, $p < .05$, findings opposite to those based on PR on L_1 and opposite to those based on both PR and NR on L_2.

Contrary to the second hypothesis, reinforcement magnitude was not significant, F (2, 84) = 1.19, $p > .05$. Furthermore, neither the four-factor interaction nor any of the two-factor or three-factor interactions including the reinforcement magnitude variable were significant.

Consistent with the third hypothesis, response force on L_2 (a response unrelated to the reinforcement contingency of the frustrated response) showed a significant increase from pre- to postfrustration trials. This effect is illustrated in Figure 1 and was presented in the context of the results with respect to the first hypothesis. It should be noted, however, that the mean force of response on L_2 (M = 4.76) was significantly greater than the mean force of response on L_1 (M = 3.68), F (1, 84) = 42.26, $p < .001$—results that might be expected since reinforcement was contingent on *S*s' restricting their force of response on L_1 within a relatively light force band, whereas reinforcement never followed L_2 responses, regardless of the force *S*s exerted.

Consistent with expectations based upon previous research (Otis, 1973), the FE was found to dissipate rapidly over time. Analyses of changes in response force from prefrustration trials to delayed postfrustration trials (Extinction Trials 3–5 on L_1 and Extinction Trials 2–4 on L_2) showed the relevant main effects and interactions to be nonsignificant.

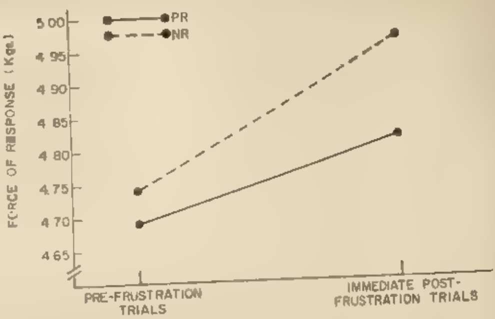


FIGURE 1. The mean Lever 2 response force as a function of trials and reinforcement type (PR = positive reinforcement, NR = negative reinforcement).

Table 2 presents the mean number of incorrect L_1 responses as a function of trial blocks for each reinforcement-type group. A Trial Blocks × Reinforcement Type analysis of variance was computed and indicated no difference between reinforcement type groups, F (1, 88) = .12, $p > .05$; a significant trial blocks factor, F (6, 528) = 22.13, $p < .001$; and a nonsignificant interaction, F (6, 528) = 1.54, $p > .05$. The significant trial blocks factor reflects the complete elimination of incorrect L_1 responses in Trial Block 7, consistent with the fact that training trials were terminated when *S*s reached a criterion of five consecutive correct L_1 responses.

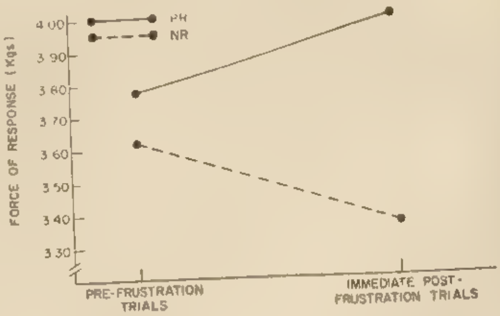


FIGURE 2. The mean Lever 1 response force as a function of trials and reinforcement type (PR = positive reinforcement, NR = negative reinforcement).

TABLE 2

Mean Number of Incorrect Lever 1 Responses during Training as a Function of Trial Blocks (Each Trial Block Represents One-Seventh of the Total Number of Training Trials for Each S) for Each Reinforcement-Type

Group	Trial blocks 1	2	3	4	5	6	7
PR							
M	1.22	1.22	1.11	1.04	.98	.71	0.0
SD	1.49	2.20	2.05	1.94	.94	.94	0.0
NR							
M	1.13	1.00	1.02	1.27	1.42	.87	0.0
SD	1.36	1.85	1.98	2.25	1.16	1.14	0.0

Note. Abbreviations: PR = positive reinforcement, NR = negative reinforcement.

Discussion

With respect to the primary purpose of the present study, the data indicate that the termination of negative reinforcement, as well as positive reinforcement, results in the FE in human *Ss*. The evidence that indicates an increase in force of response following the termination of positive reinforcement supports Amsel's (1958) frustration–drive hypothesis and is consistent with the findings of Blixt and Ley (1969). The demonstration of the FE in terms of an increase in force of response following the termination of negative reinforcement is unique with respect to both human and animal studies. With respect to other measures of the FE (e.g., start speeds from the first goal box of the double runway), the results are consistent with those of Graham and Longstreth (1970), who obtained the FE in rats using escape from shock as the negative reinforcer. It is important to note, however, that under conditions of negative reinforcement, the FE (an increase in force of response) was limited to L_2, the lever which was not instrumental in obtaining reinforcement during any phase of the experiment. With respect to L_1, the results were directly opposite of those expected, i.e., under conditions of negative reinforcement, force of L_1 responses decreased from prefrustration to postfrustration trials. This decrease may have been the result of greater aversiveness of nonreinforcement following negative reinforcement than nonreinforcement following positive reinforcement. In the present study, nonreinforcement for the negative-reinforcement groups consisted of punishment (i.e., the presentation of an aversive stimulus following the correct response); whereas nonreinforcement for the positive-reinforcement groups did not consist of punishment. If punishment leads to a depression in performance, perhaps the greater aversiveness of nonreinforcement under the NR condition resulted in a depression of response force which overrode the energizing effects of frustration. This explanation does not contradict those data that demonstrate the FE for both positive and negative reinforcement on L_2, since reinforcement and subsequent extinction contingencies were limited to L_1 responses (L_2 responses were never reinforced). This explanation proposes that under NR conditions there may be two processes which influence changes in L_1 response force following nonreinforcement: (*a*) the effects of frustration, which energize performance, thus resulting in an increase in L_1 response force, and (*b*) the effects of the aversiveness of punishment, which depress performance, thus resulting in a decrease in L_1 response force. Consistent with this explanation, the data of Table 1 show that the changes in L_1 response force from prefrustration trials to postfrustration trials were (*a*) a sharp decrease in force by Group NR-0¢, the group for which the effects of frustration would be expected to be weakest, (*b*) a slight decrease in force by Group NR-5¢, and (*c*) a slight increase in force by Group NR-10¢, the group for which the effects of frustration would be expected to be strongest.

With respect to the second purpose of the present study, both the L_1 data and the L_2 data showed a trend in the direction of a positive relationship between increases in force of response, following reinforcement termination, and magnitude of reinforcement. A perusal of the L_2 data of Table 1 indicates that the 10¢ groups showed an increase in mean response force from prefrustration trials to postfrustration trials of .18 kg., while both the 5¢ and the 0¢ groups showed increases of .15 kg. With respect to the L_1 data, the 10¢ groups showed an increase in mean response force from prefrustration trials to postfrustration trials of .14 kg., the 5¢ groups showed an increase of .04 kg., and the 0¢ groups showed a decrease of .36 kg. This trend suggests that greater manipulations of reinforcement magnitude may produce results consistent with those predicted from Spence's (1960) theory.

With respect to the tertiary purpose, the results of the present study support a generalized-drive explanation of the FE. The FE

was clearly evidenced on L_2 (the nonreinforced lever) under conditions of both positive and negative reinforcement. In addition, . . results are consistent with those of ...ne and Loesch (1967), who demonstrated ...he generalization of the FE following termination of positive reinforcement. Furthermore, the results of the present study extend the findings of Levine and Loesch to human behavior and to situations involving negative reinforcement.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

Amsel, A., & Hancock, W. Motivational properties of frustration: II. Relation of frustration effect to antedating goal factors. *Journal of Experimental Psychology*, 1957, **53**, 126–131.

Amsel, A., & Rousel, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivation complex. *Journal of Experimental Psychology*, 1952, **43**, 363–368.

Blixt, S., & Ley, R. Force-contingent reinforcement in instrumental conditioning and extinction in children: A test of the frustration-drive hypothesis. *Journal of Comparative and Physiological Psychology*, 1969, **69**, 267–272.

Daly, H. B. Evidence for frustration during discrimination learning. *Journal of Experimental Psychology*, 1971, **88**, 205–215.

Graham, T. M., & Longstreth, L. E. The role of frustrative nonreward in instrumental escape conditioning. *Journal of General Psychology*, 1970, **83**, 185–192.

Hanner, C., & Brown, P. Clarification of the instigation to action concept in the frustration-aggression hypothesis. *Journal of Abnormal and Social Psychology*, 1955, **51**, 204–206.

Lambert, J. V., & Hammond, C. J. Effect of frustrative nonrelief upon shock escape behavior in the double runway. *Journal of Experimental Psychology*, 1970, **85**, 216–219.

Levine, G., & Loesch, R. Generality of response intensity following nonreinforcement. *Journal of Experimental Psychology*, 1967, **75**, 97–102.

Marzocco, F. N. Frustration effect as a function of drive level, habit strength and distribution of trials during extinction. Unpublished doctoral dissertation. State University of Iowa, 1951. Cited by J. S. Brown, *The motivation of behavior.* New York: McGraw-Hill, 1961, pp. 208–210.

Otis, J. The effects of response delay interval and instructions on the magnitude of force of response in a nonreinforcement induced frustration paradigm. Unpublished doctoral dissertation, State University of New York at Albany, 1973.

Penny, R. K. The effects of non-reinforcement on response strength as a function of number of previous reinforcements. *Canadian Journal of Psychology*, 1960, **14**, 206–215.

Ryan, T. J., & Watson, P. Frustrative nonreward theory applied to children's behavior. *Psychological Bulletin*, 1968, **69**, 111–125.

Spence, K. W. *Behavior theory and learning.* Englewood Cliffs, N.J.: Prentice-Hall, 1960.

(Received August 11, 1974)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 824–829

THE MODALITY EFFECT: IS PRECATEGORICAL ACOUSTIC STORAGE RESPONSIBLE? [1]

RANDALL W. ENGLE [2]

King College, Bristol, Tennessee

Auditory presentation typically gives better performance than visual presentation in short-term memory experiments. This modality effect is theorized to be due to a precategorical acoustic store (PAS). Two experiments were conducted using the suffix procedure with free or serial recall of words. The findings did not support the PAS interpretation of modality effects as it now stands. That is, the suffix did not eliminate the modality effect which it should have according to the PAS interpretation.

The question is a very old one. Does auditory presentation of materials to be learned or remembered result in better learning or memory than those materials presented visually, or vice versa? A corresponding question that has attracted considerable interest is whether some people have primarily auditory memories while others favor the visual modality. This paper is concerned primarily with the initial question. If uncertainty is aroused by the second question then the reader is referred to Jensen (1971) for a partial answer.

A great deal of evidence has been compiled recently that modality of presentation is a potent variable in short-term memory experiments. The usual finding is that retention in immediate memory is much better for auditorily presented materials than for those presented visually. There is also evidence that visual presentation with concomitant vocalization of the material leads to better performance than just visual presentation (Murray, 1966). The supremacy of auditory presentation, known as the modality effect, has been observed using paired-associates (Murdock, 1966) and free recall (Craik, 1969) as well as paradigms involving order information (Murdock, 1967). The modality effect is typically obtained with both recall and recognition type retention tests.

In single-trial free recall situations the auditory superiority is most prominent over the last four or five serial input positions, in other words, the recency items. The earlier list items seem to be affected very little by modality manipulations except that occasionally a slight facilitation is found for the visual group on the primacy items (Craik, 1969; Murdock, 1966).

Craik (1969) and Murdock and Walker (1969) have attempted to delineate the theoretical implications for modality effects in short-term memory. One explanation for the results is that the short-term store is primarily auditory or acoustic in nature and auditorily presented materials enter the STS directly but visually presented materials must be translated into the acoustic code. This notion would predict auditory supremacy for all serial input positions and not, as is typically found, just those at the terminal end of the input list. Therefore, this idea seems to be somewhat discounted.

Another explanation makes use of preperceptual or prelinguistic stores. This explanation takes two forms. Crowder and Morton (1969) and Crowder (1970) have proposed that there is a precategorical acoustic store (PAS) similar in nature to the visual icon (Neisser, 1967) except of a longer duration. They argue that PAS is capable of holding information sufficiently long to affect the immediate memory task (on the order of 2 sec.). Information is supposedly lost from PAS due to displacement by subsequent acoustic events and/or decay with the passage of time. Crowder

[1] The experiments were conducted as part of a senior seminar by Glen Camenish, Luke Houchins, Kathy Lane, Terri Macray, Will Kniseley, John Saavedra and Sandy Sharp.

[2] Requests for reprints should be sent to Randall Engle, Department of Psychology, King College, Bristol, Tennessee 37620.

posits that presence of this prelinguistic information for auditory presentation is responsible for the modality effect. Crowder and Morton see no need for separate long-term and short-term memory stores. They feel that the concept of a PAS coupled with some form of articulatory coding makes the notion of a short-term store counterproductive.

The other attempt to explain the modality effect also relies on the notion of a PAS. Craik (1969) argues that the superiority of auditory presentation over the recency portion of serial position retention curves reflects the output from a postlinguistic or postcategorical short-term store which is itself indifferent with respect to input mode but which can, in some instances, be augmented by relatively unprocessed information still present in the visual and acoustic stores (icon and PAS). The modality effect is viewed as resulting from output from the short-term store being supplemented by the information in PAS for auditory presentation. Since the icon decays too quickly to be of benefit for visually presented items, the recency effect in single-trial immediate free recall is considered to be pure output from the short-term store.

As a phenomenon, the PAS has been studied almost exclusively by something called the suffix technique. The usual procedure, perfected by Crowder and his colleagues, is to present a sequence of items (typically digits) auditorily with a redundant nonrecalled suffix following the last item. This suffix is usually something like the word "zero" spoken at the end of each list. The serial recall data typically take the form of an increase in errors over the terminal (two–three) items in the suffix group as compared to a nonsuffix control group. Theoretically, the acoustic representation of the last few items in the list is still in the PAS at the time the suffix is presented and when the suffix enters the store it overwrites or interferes with the acoustic representation already there.

Since both major theoretic notions regarding modality effects rely on the PAS concept it seemed that the suffix procedure might be a useful converging operation to tease out those factors involved in the interpretation of the modality phenomenon. The first experiment was set up to be a simple demonstration that the suffix procedure would eliminate the superiority of auditory presentation over the terminal items in single-trial immediate free recall. This did not prove to be the case and thus led to the second experiment.

Experiment I

In terms of the PAS explanation, a redundant nonrecalled word at the end of auditorily and visually presented lists should have differential effects. If the suffix eliminates the ancillary information usually available in PAS for auditory presentation then the auditory suffix group should become exactly like the visual suffix and nonsuffix groups over the terminal or recency items. In other words, the prediction of Craik (1969) and Crowder (1970) should be that the suffix would affect the visual groups little, if at all. The auditory nonsuffix group should exhibit the typical supremacy over the visual groups for the terminal items, while the auditory suffix group, as a consequence of having the extra PAS information removed, should be no different than the visual groups. This should yield a Modality × Suffix and a Modality × Suffix × Serial Position interaction effect.

Method

Materials. Twenty-four lists of 12 items each were composed from the Kucera and Francis (1967) norms. The lists were recorded for auditory presentation at a 1-sec. rate and typed on paper for visual presentation by memory drum (also at a 1-sec. rate). Each list was used in both suffix and nonsuffix conditions but never with the same *S*.

Design. Half of the *S*s received the lists via auditory presentation and half by the visual mode: The visual and auditory groups were further subdivided into within-subjects suffix and nonsuffix conditions. In the suffix condition the lists ended with the word "zero." The "zero" was not to be recalled and *S* was told to either use it as a cue to recall or ignore it altogether if he so desired. The nonsuffix lists, of course, consisted of just the 12 words without the suffix. The *S*s served in the suffix and nonsuffix conditions in separate half-hour

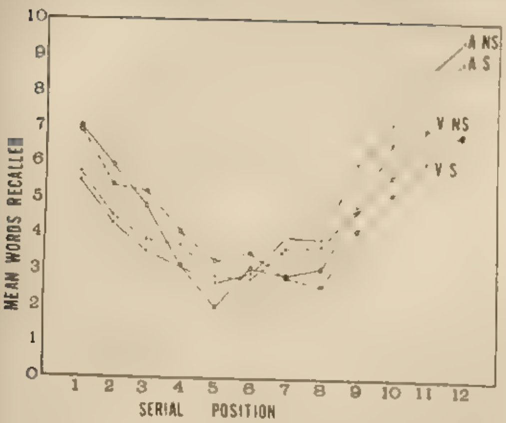


FIGURE 1. Mean free recall as a function of serial position, modality (auditory—A, or visual—V) and presence (S) or absence (NS) of a suffix.

sessions (12 lists per session) separated by 24 hr. The order of the suffix and nonsuffix sessions and set of lists in the two sessions were counterbalanced across subjects.

Subjects. The design above resulted in eight different conditions and since each of the four experimenters ran 1 *S* in each condition, 32 *Ss* were needed. The *Ss* were college sophomores at King College and served as part of a course requirement.

Procedure. The *S* was seated at a table facing either a Lafayette memory drum or Sony tape recorder depending on the appropriate condition. The *S* was instructed to recall the words in any order he deemed desirable with 60 sec. allowed for him to write his recall in a prepared booklet.

Results

The data from the first two lists in each session were discarded as practice lists. The recall data for each serial input position were combined across the remaining 10 lists for each of the four major conditions and these are depicted in Figure 1. Analysis of variance on the two modalities, two suffix conditions, and 12 serial positions showed that only the main effect of serial position and the Modality × Serial Position interaction were significant. The main effect of Serial Position resulted in F (11, 330) = 56.06, $p < .001$, and the Modality × Serial Position interaction resulted in F (11, 330) = 7.10, $p < .001$. Neither the Modality × Suffix nor Modality × Position × Suffix interactions were significant, contrary to the predictions of the PAS interpretation of the modality effect. Both interactions resulted in $F < 1$.

The absence of a Modality main effect, F (1, 30) = 1.36, $p > .10$, seemed to be due to the nearly perfect crossover of the auditory and visual groups, as is clearly shown in Figure 1. Therefore separate analyses were carried out for Positions 1–4 and 9–12. The analysis on the primacy items (1–4) resulted in significant Position, F (3, 90) = 32.38, $p < .001$, and Modality, F (1, 30) = 5.06, $p < .01$, main effects. This indicates that, for the four primacy positions, the visual presentation was, in fact, superior to the auditory presentation.

Of more importance to the present study was the analysis of the terminal positions (9–12). Auditory presentation was superior to visual presentation, F (1, 30) = 8.58, $p < .001$, showing the usual modality effect for the recency items. The Suffix and Serial Position main effects were also significant, with F (1, 30) = 5.05, $p < .01$, and F (3, 90) = 54.49, $p < .001$, respectively. None of the interactions were significant including the predicted Modality × Suffix interaction which resulted, again, in $F < 1$. This means that auditory presentation was superior to visual presentation for all four terminal positions and that the suffix did hinder recall for these terminal items. But it also means that the suffix was equally effective on these terminal items for both auditory and visual groups.

A further post hoc analysis was conducted in an attempt to ascertain the locus

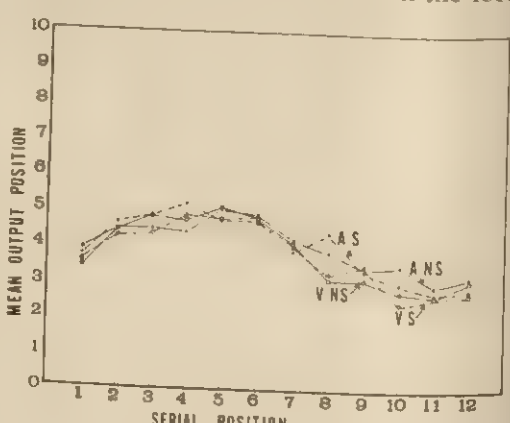


FIGURE 2. Mean output position as a function of serial position, modality (auditory—A, or visual—V) and presence (S) or absence (NS) of a suffix.

the rather surprising experimental re-ts. Following the lead of Murdock and ker (1969) the output order of the re-d items was analyzed. Figure 2 shows mean output position as a function of he recalled items input position. As can e seen from this Figure there are no major between-groups differences.

Discussion

Let us first dispense with the superiority of the visual groups to the auditory groups for the primacy items. As was mentioned in the introduction to this paper, this is an occasional result of modality experiments. Murdock (1966) used four or six paired-associates with e of the pairs being tested for retention. For his six item lists he found superiority for the visual presentation on the primacy items. Craik (1969), among others, has obtained this result with free recall, i.e., better performance on positions 1–3 for visual presentation than auditory presentation. Jensen (1971) has also reported data in which visual presentation is superior to auditory presentation. He used the memory span procedure and presented digits either visually or auditorily. The *Ss* recalled either immediately after presentation or after a delay of 10 sec. filled with a rehearsal preventative task. Jensen found that, with immediate recall, auditory presentation was slightly superior to visual presentation. However, the delayed recall results exhibited much better performance for the visual group than for the auditory group.

A probable explanation is that visual superiority in these situations is due to differential cumulative rehearsal strategies. Waugh (1960) has suggested that digit series are rehearsed cumulatively and Rundis and Atkinson (1970) have demonstrated that *Ss* in a free recall task do rehearse in a cumulative manner. Corballis (1966) found that with digit strings presented visually *Ss* did cumulatively rehearse but with auditorily presented strings there was very little cumulative rehearsal. Many subjects report a kind of "echo box" phenomenon with auditory presentation which would fit the PAS model. This quite likely leads the *S* in auditory groups to rehearse less or at least differently than the *S* in a visual group. The auditory *S* would be a great deal more passive and rely on the "echo box" while the visual *S* would need to rehearse in some manner. This would explain the differences for the primacy positions. This would also explain Jensen's (1971) findings since the PAS ("echo box") would decay leaving the auditory *S* with very little remaining information. The visual *S*, on the other hand, would be better off on delayed recall since he rehearsed the items. Other possible support for the rehearsal differences notion come from the fact that in free recall situations, facilitated primacy for visual items is usually observed when the subject serves in only the visual group or only the auditory group, i.e., between-*Ss* designs, and is not observed in within-*Ss* designs. See Craik's (1969) two experiments as evidence for this informal observation.

With respect to the analysis of the terminal positions, the prediction of the PAS explanation of modality effects seemed to be that the suffix should have eliminated any superiority for the auditory group over the visual group. The suffix did hinder recall but it did so for both the auditory and visual groups. And the suffix did not bring the auditory group down to the level of the visual nonsuffix group. This clearly does not support the PAS interpretation. One possibility for the contradiction between present findings and prior suffix procedure results might be that prior experiments using the suffix have typically used serial recall whereas the present experiment made use of free recall.

Experiment II

Since most previous experiments using the suffix procedure to study PAS have used serial recall, a second experiment was conducted to test the possibility that type of retention test contributed to the failure to support the PAS interpretation of modality effects.

Method

The same materials and design were used as in Experiment I with the exception that *Ss* were instructed to recall the items in the order they were presented. The *Ss* were 32 introductory psychology students who served as part of a course requirement.

Results

Again, the first two lists in each session were discarded as practice lists. The serial position curves are shown in Figure 3. The overall analysis of variance resulted in significant effects of Serial Position, F (11, 330) = 18.28, $p < .01$, Modality × Suffix, F (1, 30) = 12.76, $p < .01$, and Mo-

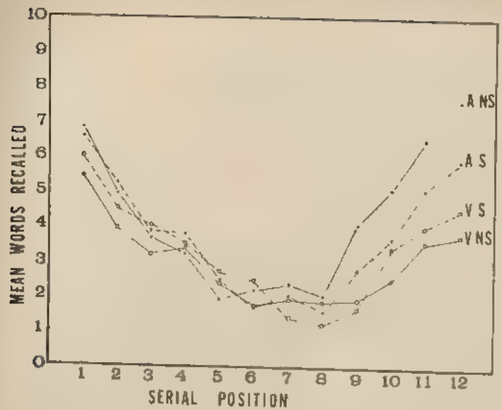


FIGURE 3. Mean serial recall as a function of serial position, modality (auditory—A, or visual —V) and presence (S) or absence (NS) of a suffix.

dality × Serial Position, F (11, 330) = 1.78, $p < .05$. Since the thrust of the study was aimed at the terminal positions, a further analysis was performed on Positions 9–12. This analysis resulted in a significant effect of Modality, F (1, 30) = 5.36, $p < .05$, showing that, for the terminal positions, auditory presentation was superior to visual presentation. The main effect of Suffix was not significant, F (1, 30) = 3.84, $p > .05$, but the Modality × Suffix interaction was significant, F (1, 30) = 13.77, $p < .01$. Paired comparisons showed the auditory suffix group to be lower than the auditory nonsuffix group, t (30) = 2.40, $p < .05$, but still significantly higher than the visual nonsuffix group, t (30) = 2.29, $p < .05$. So the suffix was effective, as the PAS interpretation predicts, in decreasing recall for auditorily presented terminal items. However, it still was not so effective as to make the auditory group identical to the visual group.

Another point of interest is that inspection of Figure 3 shows the visual superiority for primacy items found in Experiment I, to be absent in the present experiment. This supports the observation that the visual supremacy over the primacy positions is usually obtained with free recall.

Discussion

The PAS interpretation of modality effects seemed to predict that the addition of a redundant nonrecalled suffix to an auditorily presented list would result in recall curves very much like those of a visually presented list. While the suffix did have derogatory effects in both Experiment I and Experiment II it did not eliminate the superiority of auditory presentation over the recency positions.

There are several problems inherent in using the present data to argue about the contribution of PAS to the modality effects. As previously mentioned nearly all previous research on the PAS has made use of serial recall of short lists of digits. The current experiments used lists of words which were 3 4 items longer than the typical s[illegible] procedure list. Even though the second ex[illegible]ent used serial recall, there is a great diffe[illegible] between serial recall of an 8 or 9-item list [illegible]dily available digits and that of a 12-ite[illegible]t of words of considerably less availabili[illegible]nd more semantic content than the 10 d[illegible]

Nevertheless, the preca[illegible]rical acoustic store, as proposed by Crowder and Morton (1969), should not be affected by the availability of the items stored therein or their semantic content. Also, the modality effect is certainly found in nonordered recall tasks. Therefore, these tasks should be amenable to study.

One alternative might be to say that the suffix is not completely effective in eliminating the trace of those items in PAS or that the PAS is of longer duration than is typically suggested. Data pertinent to this point were reported by Murdock and Walker (1969). They randomly switched from auditory to visual within a single list and observed auditory superiority over a full 10-item list. This represented five full seconds and if PAS is responsible for their findings it must be of, at least, 5-sec. duration.

The Murdock and Walker (1969) study used words and free recall, as did the present research. This raises the completely speculative possibility that PAS is physical in nature but its duration and capacity may depend on the type of material presented, i.e., phonemes, digits, words, etc., and the type of task being used. This, of course, suggests some strategic control over the PAS.

One last word of caution should be mentioned with regard to the decrease in performance for the visual suffix group on terminal items. There is no obvious reason why this could have happened and Crowder and Morton (1969) found no such effect of suffix on recall

[a] visually presented string of digits. It may, [th]erefore, be a spurious result.

REFERENCES

[C]ORBALLIS, M. C. Rehearsal and decay in immediate recall of visually and aurally presented items. *Canadian Journal of Psychology*, 1966, **20**, 43–51.

CRAIK, F. I. M. Modality effects in short-term store. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 658–664.

CROWDER, R. G. The role of one's own voice in immediate memory. *Cognitive Psychology*, 1970, **1**, 157–178.

CROWDER, R. G. & MORTON, J. Precategorical acoustic store (PAS). *Perception and Psychophysics*, 1969, **5**, 365–373.

J[ENSEN], A. R. Individual differences in visual and [au]ditory memory. *Journal of Educational Psychology*, 1971, **62**, 123–131.

KUCERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English.* Providence, R. I.: Brown University Press, 1967.

MURDOCK, B. B., JR. Visual and auditory stores in short-term memory. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 206–211.

MURDOCK, B. B., JR. Auditory and visual stores in short-term memory. *Acta Psychologica*, 1967, **27**, 316–324.

MURDOCK, B. B., JR., & WALKER, K. D. Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 665–676.

MURRAY, D. J. Vocalization-at-presentation and immediate recall, with varying recall methods. *Quarterly Journal of Experimental Psychology*, 1966, **18**, 9–18.

NEISSER, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

RUNDIS, D., & ATKINSON, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 99–105.

WAUGH, N. C. Serial position and the memory span. *American Journal of Psychology*, 1960, **73**, 68–79.

(Received August 13, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 830–835

RECOGNITION TIME FOR WORDS IN SHORT-TERM, LONG-TERM, OR BOTH MEMORY STORES[1]

RICHARD C. MOHS AND RICHARD C. ATKINSON[2]

Stanford University

The *S*s memorized a long-term set (LT-set) of 24 words before the start of a recognition memory test. On each trial a new short-term set (ST-set) of two or four words was given; on some trials either half or all of the ST-set words were also members of the LT-set. The *S*s gave a positive response to test words that were: (*a*) only in the ST-set (new ST-set words), (*b*) only in the LT-set (LT-set words), or (*c*) in both sets (old ST-set words). Negative responses were given to words not in either set (distractors). Reaction time (RT) to both old and new ST-set words was a function of ST-set size; the RT functions for these two conditions had equal slopes but the function for old ST-set words had a smaller intercept. The ST-set size had no effect on RT to LT-set words or to distractors. In contrast to the results usually obtained using this paradigm, RT to distractors was less than RT to LT-set words. The results are interpreted in terms of a model which assumes that *S*s assess the "familiarity" of a test word before searching the memory sets.

A paradigm developed by Sternberg (1967, 1969) has been used extensively in recent years to investigate search processes in short-term memory (STM). In this paradigm *S*s are given a series of trials; on each trial *S*'s task is to decide as quickly as possible whether or not a test stimulus (probe) is a member of a previously memorized set of items (the positive set). In studies using small positive sets (six items or less) Sternberg typically finds that reaction time (RT) is a linear increasing function of the size of the positive set and that the slopes for positive and negative response functions are nearly equal. To account for these results Sternberg has proposed a theory of high speed scanning in memory. According to this theory the time between stimulus presentation and response execution is occupied by a series of stages including a stimulus encoding stage, an exhaustive serial comparison process, and a response selection stage. Because the comparison process is assumed to be serial and exhaustive, this model predicts a linear increase in RT with the size of the positive set and equal slopes for positive and negative response functions.

Although the serial–exhaustive scanning model has been supported by a number of short-term recognition experiments, data inconsistent with the predictions of the model have been reported by several investigators (see Nickerson, 1972). In an attempt to account for some of these inconsistencies Atkinson and Juola (1973, 1974) have proposed that *S*s may, in some cases, be able to execute fast responses based on a test item's "familiarity" without scanning the positive set. The familiarity model was originally formulated (Atkinson & Juola, 1973) to account for RT and error data from recognition memory studies using long, well-memorized lists of items stored in long-term memory (LTM); it has subsequently been applied both to short-term recognition studies and to studies involving searches of both STM and LTM (Atkinson, Herrmann, & Wescourt, 1974; Atkinson & Juola, 1974). The model assumes that test items (e.g., digits, words) are represented in memory as nodes and that each node, when accessed, can yield an index of that item's familiarity. Familiarity values are assumed to range over a continuum of values with the familiarity of a particular node being a function of the time since that node was last accessed

[1] This research was supported by Grant MH-21747 from the National Institute of Mental Health and Grant NSFGJ-443X3 from the National Science Foundation.

[2] Requests for reprints should be sent to Richard C. Atkinson, Department of Psychology, Stanford University, Stanford, California 94305.

relative to the number of times that node has been accessed in the past. That is, a single presentation of an infrequent item increases that item's familiarity more than does a single presentation of a frequently occurring item. Subsets of the total set of test items are defined in terms of the frequency and recency of presentation of items in that subset; for each subset there is a density function which reflects the probability that a member of that subset will yield a particular familiarity value. It is usually assumed that all familiarity distributions are normal and have a common variance but that the mean of a distribution increases with frequency and recency of presentation.

If the pool of test items used in a recognition memory test is large enough so that stimuli not in the positive set (distractors) are repeated infrequently, the mean familiarity value for items in the positive set (target items) may be considerably higher than the mean familiarity value for distractor items. In such a situation *S*s may adopt a strategy involving a check on each test item's familiarity value before the test item is scanned against the positive set. If the test item's familiarity value falls above an upper criterion or below a lower criterion, *S* executes a positive or a negative response, respectively, without a memory scan. Only if the test item's familiarity value falls between these two criteria does *S* conduct a search of the positive set. The slope and intercept of the RT function, then, are predicted to depend not only on the encoding, comparison, and response processes, but also on the proportion of trials on which *S*s execute fast familiarity responses. This proportion, in turn, depends on the distributions of familiarity values for targets and distractors and the position of the decision criteria relative to these familiarity distributions.

The present experiment was designed to investigate recognition memory processes in a task in which the items to be recognized were stored in LTM, in STM, or in both memory stores. The *S*s were given a long-term memory set (LT-set) of 24 words to be memorized before the start of the recognition test and were given an additional short-term memory set (ST-set) of 2 or 4 words at the start of each test trial. The *S*s' task was to give a positive response to test words from the LT-set or the current ST-set and to give a negative response to test words not in either set. Previous experiments (Mohs, Wescourt, & Atkinson, 1973; Wescourt & Atkinson, 1973) using this paradigm have shown that RT to test items drawn from the ST-set is a function of ST-set size, but that RT to items drawn from the LT-set and to distractor test stimuli does not depend on ST-set size. The present experiment differed from these previous experiments in that some of the words used in the ST-sets appeared on more than one trial and in more than one memory set; on some trials in the present experiment either all or half of the ST-set words were also members of the LT-set. Thus, there were three types of target words in this experiment: (*a*) words in the LT-set but not in the current ST-set, (*b*) words in the current ST-set but not in the LT-set (new ST-set words), and (*c*) words in both memory sets (old ST-set words).

Of primary concern in the present experiment was the relationship between the RT functions for positive responses to new and old ST-set words. For a given set of decision criteria the familiarity model predicts that the proportion of correct fast familiarity responses for a subset of items from the positive set varies directly with the mean familiarity value of that subset. If words in both STM and LTM at the time of test have a higher mean familiarity than do words in STM alone, the model predicts both a smaller intercept and a smaller slope for old ST-set words than for new ST-set words; that is, the model predicts that *S*s would conduct memory scans on fewer of the trials with old ST-set probes than on trials with new ST-set probes.

Method

Subjects. Eight female students at Stanford University served as *S*s. Each *S* was paid $5 for participating in two test sessions.

Stimuli. Words for the LT-set, new ST-set words, and words used as distractor test stimuli were taken from the Thorndike–Lorge (1944) word list. All words were four to eight letters in length, had two syllables, and had frequency counts of more than 20 per million. New ST-set words and distractor words were each used on only one trial in the entire experiment; that is, they were presented only once to each *S*. A single list of 24 words was used as the LT-set for all *S*s. The first 2 and the last 2 words on the list were the same for all *S*s and were tested only on warm-up trials. The middle 20 words on the list were given in a different random order to each *S*.

Ten blocks of 28 trials each were constructed. Two of the blocks served as practice blocks and were given at the beginning of the first test session; these two blocks were constructed in the same manner as were the eight experimental blocks. Within each block, ST-sets of two and four words appeared equally often. For each of the two ST-set sizes each block contained 4 trials on which the test word was a member of the ST-set: 1 trial on which the probe was a new ST-set word drawn from a pure ST-set (i.e., an ST-set containing only new words), 1 trial on which the probe was an old ST-set word drawn from a pure ST-set (i.e., an ST-set containing only old words), 1 trial on which the probe was a new ST-set word drawn from a mixed ST-set (i.e., an ST-set containing half new and half old words), and 1 trial on which the probe was an old ST-set word drawn from a mixed ST-set. For each ST-set size each block contained 3 trials on which the probe was a member of the LT-set but not a member of the ST-set: 1 trial on which the ST-set contained all new words, 1 trial on which the ST-set contained all old words, and 1 trial on which the ST-set was mixed. Negative test trials were constructed so that negative and positive responses were required equally often for each ST-set size and type (pure or mixed). Within each block the test trials were presented in a different random order.

Within each block each of the middle 20 LT-set words appeared either in an ST-set or as a test probe at least twice and not more than three times. Over all 10 blocks of trials each of the middle 20 LT-set words was tested twice as an old ST-set word and three times as a long-term memory item (i.e., on a trial when it was not a member of the ST-set). No LT-set word was tested more than once in a single block. For mixed ST-sets the position of the old words was balanced so that they appeared equally often in each position for each trial type. Also, the serial position of test words drawn from ST-sets was counterbalanced across blocks for each trial type.

Two additional blocks of six trials each were constructed for use as warm-up trials. One block was given at the start of each session. These trials were similar to those used in the regular test blocks except that one of the first two and one of the last two LT-set words were tested in each warm-up block.

Apparatus. The ST-sets for each tri[illegible]ere presented through the speaker system of a ta[illegible]ecorder.

Each test stimulus was typed in ca[illegible]l letters on a white 6 × 9 in. index card wit[illegible]n IBM Executive Registry typewriter. The s[illegible]uli were presented in an Iconix tachistoscope[illegible]ey appeared at the center of the visual fi[illegible]d subtended a visual angle of about 2 degree[illegible]fixation target of four dots in the shape of a r[illegible]gle was displayed at the center of the field fo[illegible]c. prior to the onset of the test stimulus. Th[illegible]ay was dark between trials.

On a table to the right of *S* three [illegible]buttons were arranged along an arc. The *S* co[illegible]est her forefinger on the center button between[illegible]ls and make short movements to the left or ri[illegible] strike one of the two response buttons. Hal[illegible]the *S*s were assigned the right button as the p[illegible]tive response and the left button as the negative [illegible]sponse; the assignment was reversed for the othe[illegible]. The *S* held in her left hand a button used [illegible]nitiate display of the test stimulus on each trial

Procedure. Each *S* was contacted by[illegible]one at least 18 hr. prior to the first test sessio[illegible]The *S* was given the LT-set and was told to mem[illegible]ze the list so that it could be recalled in correct s[illegible] order at the first test session. Test sessions w[illegible]cheduled for two successive days.

All *S*s satisfied a learning criterion for th[illegible]T-set by giving both written and oral recalls of th[illegible]T-set in correct serial order at the start of the f[illegible]t test session. After completing both recalls correctly *S* was seated at the tachistoscope and given instructions about the task. The *E* explained that the experiment involved a series of test trials and that the following sequence would be followed on each trial: (*a*) the *E* would start each trial by saying "ready"; (*b*) the *E* would turn on the tape recorder and *S* would hear an ST-set of either two or four words; (*c*) the *S* would then repeat the ST-set aloud; (*d*) when *S* was ready she was to push the start button in her left hand; (*e*) this would cause a fixation box to appear in the tachistoscope for .5 sec., and then a test word would appear; (*f*) the test word would remain visible until *S* made a positive or a negative response by pressing the appropriate button; (*g*) the *S* would then repeat the ST-set aloud once more (this was done to ensure that old ST-set words were maintained in STM as well as LTM during the recognition test); and (*h*) the *E* would record the RT and a new trial would begin. The *E* then explained fully the nature of the ST-sets; the *S* was told that some of the ST-sets would be all new words, that some would contain only words already in the LT-set, and that some would contain half new and half old words. The *S* was not told that only the middle 20 words of the LT-set would be tested during regular test blocks. The *S* was told that the positive set on each trial consisted of the words in the LT-set, words presented from the tape recorder on that trial, and words in both the ST-set and the LT-set. It was explained that new ST-set words and distractor words would

...ed on only one trial in the entire experiment. ...nstructions emphasized that S was to respond ...dly as possible while trying to avoid errors. ...E was sure that S understood the instructions, ...first warm-up block was given. After the ...rm-up block S was asked if she had any questions. When S's questions, if any, were answered, the two practice blocks and three test blocks were given. During testing Ss were informed whenever they had made an error. The time between trials was about 10 sec.

At the start of the second day's session S was again required to give both a written and an oral recall of the LT-set. The instructions were then reviewed and S was given another warm-up block and five regular test blocks. Each session lasted about 1 hr.

Results and Discussion

The RT data from eight test blocks, the last three given in the first session and the five given in the second session, were included in the analysis. Data from trials on which an incorrect response was made were excluded from the analysis.

A preliminary analysis of the RT data for positive responses to words in the ST-set suggested that ST-set type (pure vs. mixed) had no effect on RT. This variable was included in the experiment in order to determine whether the familiarity value of a test word from the ST-set was affected by the familiarity of other ST-set words. An analysis of variance using S means as scores was performed with ST-set type, ST-set size (two or four words), and probe type (new or old) as factors. The analysis showed that the main effect of ST-set type and all interactions involving ST-set type were not significant, F (1, 7) < 1 in all cases.

Since ST-set type clearly had no effect on RT, the data for pure and mixed ST-sets were pooled. Table 1 presents a summary of the pooled data. Trials on which Ss gave a positive response to a new ST-set word are labeled Y(ST-new); trials on which Ss gave a positive response to an old ST-set word are labeled Y(ST-old); trials on which Ss gave a positive response to an LT-set word not in the ST-set are labeled Y(LT); and trials on which Ss gave a negative response to a distractor word are labeled N(D). In the second column of Table 1 are the two ST-set sizes for each condition, and in the third column is the total number of observations made for all Ss over eight test blocks. Using the raw data collapsed over ST-set types, the mean RT for each of the eight trial types was computed for each S. Outlying scores for each S were eliminated from computations by the following method: (*a*) for each trial type, the mean was calculated, (*b*) each score that was more than 1.5 times the mean was deleted, and (*c*) the mean was recalculated from the remaining scores. This procedure eliminated about 2.5% of the scores. The mean and standard deviation of S means for each of the trial types are presented in the fourth and fifth columns of Table 1. The last column presents the error rate for each trial type.

TABLE 1

Number of Observations, Reaction Time Means, Standard Deviations, and Error Rates for All Trial Types

Condition	ST-set size	No. of observations	*M* (msec.)	*SD* (msec.)	Percent errors
Y(ST-new)	2	128	626	133	1.6
	4	128	663	122	.8
Y(ST-old)	2	128	604	111	.0
	4	128	636	108	.0
Y(LT)	2	192	766	139	3.7
	4	192	771	108	8.6
N(D)	2	448	752	158	.9
	4	448	762	184	.2

A graph of the means from Table 1 is presented in Figure 1. Lines for the Y(LT) and N(D) conditions were drawn at the mean RT for the two ST-set sizes while the lines for the Y(ST-new) and Y(ST-old) conditions connect the two points for each condition. An equation for each line is given in the figure.

A comparison of the RT functions for the Y(ST-old) and Y(ST-new) conditions presented in Figure 1 indicates that the intercept for the Y(ST-old) condition is slightly smaller than the intercept for the Y(ST-new) condition but that the slopes for the two conditions are nearly identical. A two-factor analysis of variance using individual S means for these two condi-

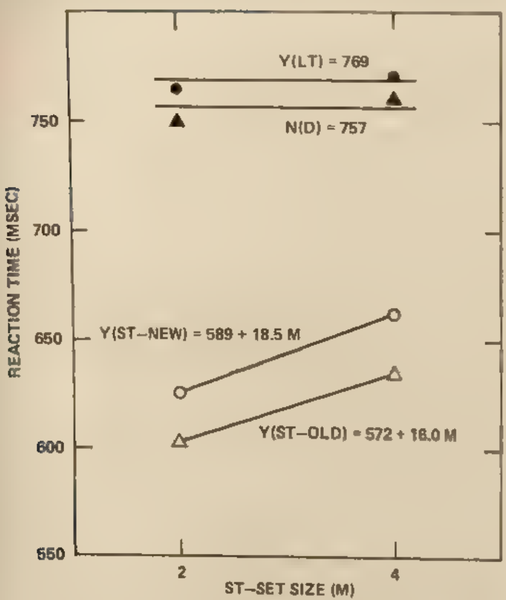


FIGURE 1. Mean reaction time vs. short-term (ST) set size for all conditions. (See text for explanation of conditions.)

tions as scores showed that the effect of ST-set size on RT was significant, F (1, 7) = 10.7, $p < .025$, that the effect of probe type was marginally significant, F (1, 7) = 4.1, $.05 < p < .10$, and that the interaction between these two variables was not significant, F (1, 7) < 1. The lack of a significant interaction between set size and probe type suggests that the difference in RT between new and old ST-set words was not due to a difference in familiarity; a greater mean familiarity for old ST-set words would have produced both a smaller slope and a smaller intercept for the Y(ST-old) function. It is possible, however, that the intercept difference shown in Figure 1 reflects the fact that old ST-set words were encoded more quickly than were new ST-set words. That recognition involves a stimulus encoding process which is separate from the comparison process has been demonstrated by Sternberg (1967). Furthermore, this interpretation of the intercept difference is consistent with the results of previous experiments in which familiarity could play no role (Kreuger, 1970; LaBerge & Tweedy, 1964), showing that RT decreases for more frequent test stimuli even when response frequency is controlled. An encoding difference could have resulted, then, because old ST-set words were each used as probes on several trials while new ST-set words were each presented only once in the entire experiment.

The fact that no evidence was found indicating a greater mean familiarity for old ST-set words than for new ST-set words suggests that a limit on the range of mean familiarity values must be included in the familiarity model. That is, the distribution of familiarity values for items currently being maintained in STM has a mean that is as great, or nearly as great, as the mean familiarity value for any set of items in LTM, STM, or both stores.

Paired t tests using individual S means as scores indicated that ST-set size had no effect on RT for either the Y(LT) condition, t (7) < 1, or the N(D) condition, t (7) < 1. Previous experiments using a similar paradigm (Mohs et al., 1973; Wescourt & Atkinson, 1973) have also found that ST-set size had an effect on RT to probes from the ST-set but had no effect on RT to probes from the LT-set or to distractor test stimuli. A model that accounts for this basic result by assuming that Ss search STM and LTM in parallel rather than sequentially is presented in both of the cited papers.

The present results are unique, however, in that the N(D) condition had a smaller RT than did the Y(LT) condition. In both of the previous studies of this type (Mohs et al., 1973; Wescourt & Atkinson, 1973) RT to distractor words was over 40 msec. greater than RT to LT-set words. The familiarity model can account for this reversal if it is assumed that Ss in the present experiment set upper and lower criteria in such a way that a high proportion of LT-set words had familiarity values between the criteria and so that a high proportion of the distractor words had familiarity values below the lower criterion. The reason for Ss' positioning the decision criteria in such a manner in the present experiment may be the following: Some Ss reported that they often decided whether a test word was a member of the ST-set,

LT-set, or both memory sets before ...g a positive response. By positioning ...e decision criteria in the manner de...ed above, *S*s could ensure that memory ...rches would be conducted on a high ...roportion of the positive trials and that high proportion of the distractor words ...ould be correctly rejected on the basis of ...amiliarity.

The error data shown in Table 1 provide ...rther support for this explanation of the ...eversal. The familiarity model predicts ...hat errors occur when the familiarity value for a member of the positive set falls below the lower criterion or when the familiarity value for a distractor falls above the upper criterion. We are assuming here three ...verlapping distributions of familiarity values, each normal and with common variance; we assume that the distribution for distractors had a low mean, that the distribution for ST-set words had a high mean, and that the mean for LT-set words was intermediate. If most of the familiarity values for LT-set words were between the criteria, few distractors would have familiarity values above the upper criterion and few ST-set words would have familiarity values below the lower criterion. Given these criteria, then, the model predicts lower error rates for ST-set words and distractors than for LT-set words.

REFERENCES

Atkinson, R. C., Herrmann, D. J., & Wescourt, K. T. Search processes in recognition memory. In R. L. Solso (Ed.), *Theories in cognitive psychology.* Potomac, Md.: Lawrence Erlbaum Associates, 1974.

Atkinson, R. C., & Juola, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Fourth International Symposium on Attention and Performance.* New York: Academic Press, 1973.

Atkinson, R. C., & Juola, J. F. Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology.* San Francisco: W. H. Freeman, 1974.

Kreuger, L. E. Effect of stimulus probability on two-choice reaction time. *Journal of Experimental Psychology*, 1970, **84**, 377–379.

LaBerge, D., & Tweedy, J. R. Presentation probability and choice time. *Journal of Experimental Psychology*, 1964, **68**, 477–481.

Mohs, R. C., Wescourt, K. T., & Atkinson, R. C. Effects of short-term memory contents on short- and long-term memory searches. *Memory & Cognition*, 1973, **1**, 443–448.

Nickerson, R. S. Binary-classification reaction-time: A review of some studies of human information-processing capabilities. *Psychonomic Monograph Supplements*, 1972, **4**(17, Whole No. 65).

Sternberg, S. Two operations in character recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53.

Sternberg, S. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

Thorndike, E. L., & Lorge, I. *The teacher's word-book of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

Wescourt, K. T., & Atkinson, R. C. Memory scanning for information in short- and long-term memory. *Journal of Experimental Psychology*, 1973, **98**, 95–101.

(Received August 16, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 836–840

PICTORIAL ELABORATION AND RECALL OF MULTILIST PAIRED ASSOCIATES[1]

EDWARD J. FORBES[2] AND HAYNE W. REESE

West Virginia University

In learning multiple paired-associate lists, all with the same stimuli, *S*s may form a new image for each pair or may progressively elaborate the image for a List 1 pair by incorporating each new response item into the image. In the latter, imagery and response-repetition effects are confounded. Analogues of these imagery types were manipulated by imposed pictorial elaboration in 4 paired-associate lists, with college *S*s. A control for response repetition was included. The elaboration types were found to be equally superior to their controls, for both serial and free recall, and for both immediate and 1-week delayed testing. Thus, elaboration effects are not lost in long-term retention, serial information is not lost in progressive elaboration, and response repetition and elaboration have separable effects.

It is now well established that nonverbal mnemonics, or images, facilitate the learning and short-term retention of a paired-associate list when the images represent interactions between the referents of the stimulus and response items (see review by Paivio, 1971). However, there is contradictory evidence about whether long-term retention is facilitated by these interactive images (e.g., Palermo, 1970).

In multilist paired-associate tasks in which the same stimuli appear with new responses in the different lists, there is some evidence that interactive images facilitate performance, perhaps by reducing associative interference between lists (e.g., Bower & Reitman, 1972; see brief review by Reese, 1972). One mnemonic strategy that *S*s sometimes report (Bugelski, 1968) is to elaborate a progressively more complex image, or "grand imaginal scene" (Bower, 1972), for a given stimulus by incorporating into the original interactive image each new response to be associated with that stimulus in the successive lists. Bower (1970) had suggested that information about serial order might be lost in these built-up images; but Bower and Reitman (1972) obtained contradictory evidence.

Bower and Reitman compared two groups, one given progressive elaboration instructions and the other instructed to form an interactive image for each pair and to replace it with a new image on each new list. Free and serial recall were better in the progressive-elaboration group than in the single-response elaboration group. Furthermore, the effects were obtained in both short-term and long-term (1 wk.) retention.

With progressive elaboration there is repetition of old-list responses in each new-list elaboration of the image. This repetition could be the source of the superiority of progressive elaboration. It would facilitate free recall by increasing the number of learning "trials," and it could facilitate serial recall by giving the response items different "frequencies" in Underwood's (1969) sense, as Reese (1970) suggested. However, if repetition is the source of facilitation, it should be equally effective without the context of imagery. A problem, then, is to determine whether repetition per se is the effective variable in progressive elaboration or whether the interactive imagery involved provides further facilitation. In the Bower and Reitman (1972) study, there were no control conditions in which imagery instructions were omitted, and consequently there was no way to separate the repetition and imagery effects.

To control for effects of repetition, Reese (1972) used imposed pictorial materials generating analogues of the imagery and

[1] A report of the study was given orally at the meeting of the Eastern Psychological Association, Boston, April 1972.

[2] Requests for reprints should be sent to Edward J. Forbes, Department of Psychology, West Virginia University, Morgantown, West Virginia 26506.

trol conditions. The analogue of giving ery instructions is to show, for each picture of the stimulus and response interacting; the analogue of omitting imagery instructions is to show the mulus and response items side by side d not interacting. In both kinds of con- l condition, S might spontaneously rm interactive images, but the proba- lity of interactive images is assumed to reduced. It was found in a retention st that progressive elaboration was su- erior to single-response elaboration, as in the Bower and Reitman (1972) study; but with response repetition controlled, reten- tion was the same in the two kinds of elaboration condition, and in fact neither was superior to its comparable control con- dition, contrary to the implications of the Bower and Reitman study. However, Reese tested recognition memory instead of recall, and the recognition test was so con- structed that serial retention could not be tested. (In addition, his *Ss* were young children.)

The present study can be considered to be an extension of the Bower and Reitman study, with the response repetition con- trols included (although this study was actually designed and run before the Bower and Reitman study was published). Be- cause of the nature of the control condi- tions, it was decided to use imposed pic- torial materials, as in the Reese (1972) study, but with recall instead of recogni- tion tests. Free and serial recall were tested immediately and again after a one- week delay.

Method

Design. The experimental manipulation involved the presence or absence of imposed pictorial elabora- tion. There were four conditions, each with four lists. (*a*) In a single-response elaboration condition, the stimulus and response items in a given pair in a given list were depicted as interacting in some way (see Reese, 1972, for details). The "scene" was shown only in the list in which it was appropriate. (*b*) In a "standard" control condition, the stimulus and response items in a pair were depicted side by side and not interacting in any way. A given pair was shown only in the list in which it was ap- propriate. This condition is comparable to the single-response elaboration condition, in that both involved presentation of only the currently appro- priate response items in each list. (*c*) In a progres- sive-elaboration condition, each new response item was added to the scene from the preceding list to form a progressively more complex scene that finally incorporated all four of the appropriate response items. The interaction between a response item and the stimulus item was exactly the same as in the single-response elaboration condition. (*d*) In a "row" control condition, the stimulus and response items in a pair in List 1 were presented side by side and not interacting, exactly as in List 1 for the "standard" control condition. In each successive list, however, each new response item was added to the row without deleting the response items already in the row. This condition is comparable to the progressive-elaboration condition, in that both in- volved presentation of the currently appropriate response items in each list together with all previ- ously appropriate response items, thus controlling the amount of sheer repetition of the response items.

The four conditions form a two-by-two factorial matrix with the dimensions: (a) Picture Treatment —imposed vs. no imposed elaboration—and (b) Response Treatment—response items repeated vs. not repeated in subsequent lists. In addition, the design included two other variables of experimental concern, both manipulated within *Ss*: (*c*) Test— free vs. serial recall—and (*d*) List. A fifth variable included in the analyses was (*e*) Order of Tests, re- flecting counterbalancing of the free and serial recall tests. The data from a sixth variable—immediate and delayed testing—were analyzed separately.

Subjects. The *Ss* were 48 students in sections of an introductory psychology course at West Virginia University, who volunteered to serve as *Ss* for extra credit in the course. They were run in groups of 2 to 11 persons. These groups were assigned to con- ditions in such a way that time of testing was varied, the sex distribution was roughly balanced, and inso- far as possible there was rotation through the con- ditions, roughly equating control or experimental conditions in number of *Ss* before repeating the rotation.

Materials. The materials were identical to those used and described by Reese (1972). They were line drawings of animals, used as stimuli, and of common objects, used as responses. They were divided into four five-pair lists, all with the same stimulus items and all with different response items. The pairings were random, except that obviously high associates were not paired. The materials were photographed for slide projection as positives. Projection time was controlled by Hunter decade timers.

Procedure. The *Ss* were run in subgroups, in each of which all *Ss* served in the same condition, but roughly half were given one sequence of the free and serial recall tests and the others were given the reverse sequence.

A study-test procedure was used, in which eight study trials and two test trials were given. Each list was presented twice for study, one pair at a time, before the next list was presented. The order of the pairs was changed in a predetermined, ran-

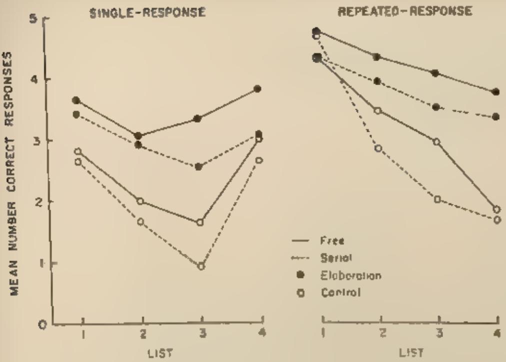


FIGURE 1. Recall in immediate tests.

dom way for each trial, with the restriction that across the eight study trials (two per list, four lists) the mean within-list positions of the stimuli were about the same for all stimuli. The presentation rate was 4 sec. per pair.

The test trials were given immediately after the eighth study trial. Free recall was tested in one of these trials and serial recall in the other. Half of the *S*s had the free-recall test first and the others had the serial-recall test first. The answer sheet for the free-recall test included a list of word labels of the stimulus items in the same predetermined random sequence for all *S*s, and beside each word was a line 5 in. long (12.7 cm). The instructions, printed at the top of the sheet, were to write down the names of the response items in any order. The answer sheet for the serial-recall test included a list of the words, in a different predetermined random sequence from that in the free-recall test to reduce transfer effects, and beside each word were four short lines. The instructions were to write down the names of the response items in the original order of appearance. The duration of each test trial was 3 min. Guessing was explicitly encouraged in the instructions for each test.

Seven days after the initial session, the *S*s were given the two test trials again, in the same sequence as in their initial test and with no further study presentations. The *S*s had not been told that these retests would be given. The *S*s were tested in groups of 15 to 20 in their classroom (these groups included *S*s participating in another study).

Results

Before the detailed findings are presented, it is useful to note that the Bower and Reitman (1972) findings were broadly replicated. In both free and serial recall, tested immediately and after the delay, the progressive-elaboration group was superior to the single-response elaboration group (all ps $< .01$). However, when these groups were compared with their respective control groups, the two types of elaboration were found to be almost exactly equal in overall effectiveness, as detailed below.

Immediate tests. Figure 1 shows the mean numbers of correct responses per list in the elaboration and control groups for free and serial recall. It can be seen, first, that both kinds of elaboration were facilitative, relative to the control conditions. In confirmation, the main effect of picture treatment was significant, F (1, 40) = 15.42, $p < .001$; and although picture treatment interacted significantly with list, F (3, 120) = 3.08, $p < .05$, follow-up t tests indicated that the difference was significant in all lists except the first one (List 1, t = 1.19; all other ts > 3.00, $p < .01$; these ts are Cochran-Cox approximations, with error terms derived from the overall analysis of variance). Second, within each response-treatment condition it can be seen that free recall was better than serial recall, but that the relative effect of elaboration (i.e., the elaboration vs. control difference) was about the same for serial recall as for free recall. In confirmation, picture treatment and test did not interact with each other ($F < 1.00$) nor jointly with any other variable (largest $F = 1.85$, for Picture Treatment × Test × Response Treatment × Order of Tests, df = 1, 40). Finally, it can be seen that the relative effects of elaboration on free and serial recall were about as strong in the repeated-response condition as in the single-response condition. In confirmation, picture treatment, test, and response treatment did not interact (all Fs involving all three variables $<$ 1.00 except that F = 1.85 for the four-way interaction mentioned above). These results show that elaboration facilitated performance, and affected serial and free recall in the same way. Furthermore, single-response elaboration was as facilitative as progressive elaboration when the effect of response repetition was controlled.

The source of facilitation in the elaboration groups seems to have been a reduction in interlist interference. It has already been noted that the elaboration and control groups did not differ significantly in

List 1, and examination of Figure 1 shows that the curves of the elaboration groups were flatter than those of the control groups. The Picture Treatment × List interaction was significant, as already mentioned, and follow-up t tests confirmed the apparent trends. In the combined elaboration groups, List 1 performance was significantly superior to List 3 and 4 performance, and no other differences between lists were significant. Thus, performance dropped after List 1 but immediately leveled off. In the combined control groups, performance was significantly better in List 1 than in all other lists, and significantly better in List 2 than in List 3. Thus, performance dropped continuously through the first three lists before leveling off.

It can also be seen in Figure 1 that typical serial position curves were obtained in the single-response groups and that there was a marked primacy effect and no recency effect in the repeated-response groups. These trends were confirmed by t tests, which were justified by a significant interaction between response treatment and list, F (3, 120) = 10.89, $p < .001$. It is, of course, not difficult to explain the primacy effect in the repeated-response conditions. The effect, which has also been obtained in a recognition test (Reese, 1972), presumably resulted from the extra practice provided when the previously relevant responses were repeated.

The only other significant effects are of no systematic interest: the main effects of response treatment (of no interest because of the interaction with list), test, order of tests, and list; and the interactions of test with order of tests and with list.

Delayed tests. The data from the delayed tests are summarized in Figure 2. (Group means were substituted for the data of one *S* who missed the delayed tests, with appropriate adjustment of *df* in the analyses of variance.) Overall recall was better in the elaboration conditions than in the control conditions, F (1, 39) = 10.98, $p < .005$, but the apparent effect of response treatment was not significant, F (1, 39) = 3.72, $p = .07$.

The only other significant effects were

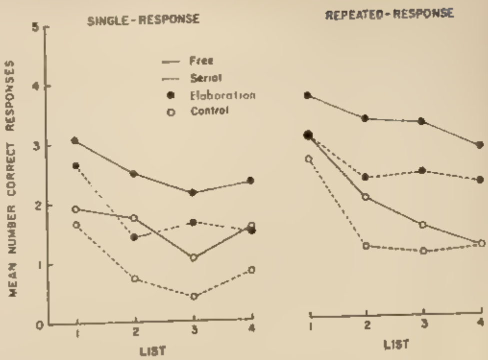


FIGURE 2. Recall in delayed tests.

the main effects of test and list (ps $< .001$) and their interaction, F (3, 117) = 3.71, $p < .025$. Follow-up t tests showed that the recency effect disappeared but the primacy effect remained, and was stronger for free recall than for serial recall. The disappearance of the recency effect wiped out the effect of response treatment on the shape of the retention curves. The effect of elaboration on the shape of the curves did not approach significance (F = 1.12).

The superiority of the elaboration conditions on the delayed tests is ambiguous because of the differences in immediate recall; more was learned, therefore more was retained. To determine whether elaboration affected long-term retention independently of amount learned, the amount of *loss* was analyzed. (The conditional probability of a correct response in a delayed test, given a correct response in the analogous immediate test, could also be used, but because of reminiscence effects it should underestimate retention.)

The only significant effect involving picture treatment was the triple interaction with list and order of tests, F (3, 117) = 3.58, $p < .025$. Of the eight elaboration vs. control differences involved in this interaction, four favored the elaboration group and four the control group, but there was no obvious pattern, overall, and only one difference was significant, favoring the elaboration group in List 1 on the test given second. Thus, in general, loss was not reliably related to picture treatment, indicating that amount lost was generally unrelated to amount originally learned.

The response treatment by list interaction was significant, $F\ (3,\ 117) = 6.95$, $p < .001$. Follow-up t tests showed that in the single-response conditions the most loss was in the recency effect, as theoretical explanations of this effect would predict; and in the repeated-response conditions the most loss was in the primacy effect.

The only other significant effects were order of tests, test, and test by list.

Discussion

Short-term retention. The results show that response repetition facilitates learning of earlier-list responses, presumably because of the extra practice available. Response repetition in the context of elaboration provides more facilitation than response repetition without elaboration, but no more than is provided by single-response elaboration (presumably without response repetition). By analogy, it can be inferred that built-up images facilitate recall partly because of the effect of interactive imagery on retention and partly because of the response-repetition effect on learning. These effects are additive, and equally strong for free and serial recall.

It was expected that free recall would be facilitated more by progressive elaboration than by single-response elaboration. However, they produced equal facilitation, when response repetition was controlled, by equally reducing interlist interference. By analogy, the imagery component of built-up images functions like that of single-response interactive images, and the function is to reduce interference. Thus, interactive imagery per se may produce a basic amount of facilitation, regardless of the complexity of the image (see also Bower, 1972).

An alternative interpretation is possible. Recoding of the stimulus items may have occurred in the single-response elaboration condition (see Keppel & Zavortink, 1969) but not in the progressive-elaboration condition. If so, then the cause of facilitation in the former condition would be reduced stimulus similarity (converting A–B A–C to A–B D–C), and the cause of facilitation in the progressive-elaboration condition might be chunking of the materials (see, e.g., Horowitz, Lampel, & Takanishi, 1969; Taylor, Josberger, & Prentice, 1970). It is possible, in other words, that the finding of equal facilitation for the two kinds of elaboration was fortuitous and does not reflect an underlying single source of facilitation. However, these particular different sources of facilitation do not account for the finding that progressive elaboration and single-response elaboration equally facilitated serial recall as well as free recall. It therefore seems more reasonable to posit a single source of facilitation in the two elaboration conditions, presumably an interactive imagery component. Why it works is not clear, but it appears to function similarly regardless of the complexity of elaboration.

Long-term retention. For long-term retention, neither kind of elaboration provided any *facilitation.* Elaboration effects appeared, but if elaboration had facilitated long-term retention, there would have been less *loss* in the elaboration conditions than in the control conditions. Thus, elaboration did not improve long-term retention; but neither did it interfere.

References

Bower, G. H. Analysis of a mnemonic device. *American Scientist*, 1970, **58**, 496–510.

Bower, G. H. Mental imagery and associative learning. In L. W. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley, 1972.

Bower, G. H., & Reitman, J. S. Mnemonic elaboration in multilist learning. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 478–485.

Bugelski, B. R. Images as mediators in one-trial paired-associate learning. II. Self-timing in successive lists. *Journal of Experimental Psychology*, 1968, **77**, 328–334.

Horowitz, L. M., Lampel, A. K., & Takanishi, R. N. The child's memory for unitized scenes. *Journal of Experimental Child Psychology*, 1969, **8**, 375–388.

Keppel, G., & Zavortink, B. Further test of the use of images as mediators. *Journal of Experimental Psychology*, 1969, **82**, 190–192.

Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.

Palermo, D. S. Imagery in children's learning: Discussion. In H. W. Reese (Chm.), Imagery in children's learning: A symposium. *Psychological Bulletin*, 1970, **73**, 415–421.

Reese, H. W. Implications of mnemonics research for cognitive theory. Paper presented at the Southeastern Conference on Research in Child Development, Athens, Georgia, April 1970.

Reese, H. W. Imagery and multiple-list paired-associate learning in young children. *Journal of Experimental Child Psychology*, 1972, **13**, 310–323.

Taylor, A. M., Josberger, M., & Prentice, J. L. Imagery organization and children's recall. Paper presented at the meeting of the American Educational Research Association, Minneapolis, March 1970.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

(Received August 20, 1973)

... mental Psychology
... No. 5, 841–844

IMAGERY AND CUED RECALL: CONCRETENESS OR CONTEXT?[1]

RONALD C. PETERSEN[2]

University of Minnesota

Two hypotheses regarding the recall of noun triplets through the use of imagery were investigated in a cued-recall study. An extension of Paivio's conceptual-peg hypothesis for paired-associate learning predicted that the concreteness of the cue word would be the critical factor in determining recall, while a constructive approach to imagery predicted that the retrieval of the context used to construct the image would be the essential element. Sixty-four *S*s formed images for 20 noun triplets with corresponding contexts and were later presented with one member of each triplet and asked to recall the other two members. Results indicated that the elicitation of the context was the single factor determining recall, while concreteness had no consistent effect. Implications of the results for the processes involved in remembering through imaging are discussed.

Imagery has been shown to be an extremely potent variable in many types of learning and memory tasks (see Bower, 1972, and Paivio, 1971, for reviews). While not a greal deal is known about the mechanism by which imagery works, several hypotheses have been offered to explain its effects. For example, to account for the facilitative effect of imagery on paired-associate learning, Paivio (1963, 1965) has offered a conceptual-peg hypothesis. This hypothesis states that when two words are presented together on a study trial, *S* forms a compound image of the referents of these words, such that when one of the words is subsequently presented to *S* on a test trial, it redintegrates the compound image from which the response term is decoded.

From this hypothesis it follows that recall should depend upon the imagery value (*I*) of the members of the pair and, in particular, upon the *I* of the stimlus term. The reasoning for this is that on recall trials, stimulus terms rated high in *I* will be more likely to redintegrate the compound image than those rated low in *I*. A substantial amount of research has been generated by this hypothesis and, almost without exception, has supported the basic position (e.g., Paivio, 1965, 1969, 1971).

Alternatively, if imagery is viewed as a constructive process (Bartlett, 1932; Neisser, 1967), the conceptual-peg hypothesis no longer appears relevant because images are not thought to be "stored" and "retrieved" as such. Instead, an important component of the constructive approach to imagery is the setting or context used to construct the image, and the critical element at the time of recall is the elicitation of this context from which the image can be reconstructed. In other words, a concrete word would be effective as a recall cue only insofar as it elicited the context in which the image was originally constructed. There is nothing inherent in the concreteness itself that makes it particularly effective, and, in fact, an abstract cue word could be more effective than a concrete word as a recall cue if it were more likely to elicit the context.

The present experiment was designed to evaluate these two views of imaging.

[1] This research is based on a PhD dissertation completed at the University of Minnesota. The work was supported in part by grants to the Center for Research in Human Learning at the University of Minnesota from the National Science Foundation (GB-17590), the National Institute of Child Health and Human Development (HD-01136 and HD-00098), and the Graduate School of the University of Minnesota. The author would like to thank his adviser, Paul W. Fox, and the other members of his committee, Raymond O. Collier, Jr., James J. Jenkins, Paul E. Johnson, and David L. Laberge, for their assistance with this research.

[2] Requests for reprints should be sent to Ronald C. Petersen, who is now at the Biomedical Laboratory, Building 3100, Edgewood Arsenal, Aberdeen Proving Ground, Maryland 21010.

TABLE 1

Sample Set of Items as They Appeared to *Ss* in the Associative Rating Task

Cue–context	Responses (high–low)			
Bride–Wedding	4	3	2	1
Glove–Wedding	4	3	2	1
Beauty–Wedding	4	3	2	1
Candle–Wedding	4	3	2	1
Love–Wedding	4	3	2	1
Vows–Wedding	4	3	2	1

Word triplets were used as the materials to be remembered, and each triplet was provided with a context in which the words could be readily imaged. The *Ss*' task was to construct an image for the three words in the setting provided by the context word, such that, when later presented with one of the items in the triplet, they would be able to recall the other two. The conceptual-peg hypothesis predicts that the concreteness of the cue word is the critical element for determining recall; whereas the constructive view contends that the essential factor for recall is the capacity of the cue word to elicit the context in which the image was constructed. These two variables were factorially manipulated to determine which is prepotent.

Preliminary Study

The primary variables in this study involved scaled attributes of the cue words. For example, suppose *Ss* were to form an image for the words, DRESS, RING, and CANDLE, in the context of WEDDING. If CANDLE were to be the cue word for recall, it would be essential to know the *I* of CANDLE and the likelihood with which it would elicit the context, WEDDING, at the time of recall. Therefore, imagery and associative norms were compiled on the materials. The noun triplets were selected by *E* such that the three words were obviously related to a common setting or context. Two of the items were common to all experimental conditions and were concrete nouns which could be readily imaged in the contexts. The third member of the triplet, the cue word, varied on the two scaled dimensions of imagery and association value. The latter dimension assessed the capacity of the cue word to elicit its context. To scale the cue word for each context, several candidates for the four-fold classification of concrete–abstract and high–low association value of cue word to context were selected by *E* for the two norming exercises.

Associative Rating Task

A particular type of associative norming procedure was developed to assess the degree to which the cue words would elicit the context words in this experiment. It was imperative that the measure reflect the likelihood with which a particular cue word (one of the members of the triplet) would elicit the particular context word used for the formation of the image of that triplet. In this sense, the present norming task was developed to reflect the processes proposed to take place at the time of recall. A list of potential stimulus terms was selected, and each term was paired with its context for presentation to a group of *Ss*.

Method

Subjects. The *Ss* were 36 male and 40 female volunteers from an introductory psychology course at the University of Minnesota. Each *S* received course credit for participating.

Procedure. The associative rating task was conducted in one session. Table 1 shows the format for the exercise in which several pairs of words (cue words with contexts) were grouped under their appropriate contexts. The *Ss* were told to interpret the pairs from the frame of reference of the context word which preceded each group of pairs. The *Ss*' task was to estimate for each pair what proportion of a large group of people would respond with the right-hand member of each pair (context) when presented the left member (cue word). The *Ss* made their responses by circling a number from 1 through 4, where 1 indicated that a low proportion and 4 indicated that a high proportion of the population would give the member on the right when presented with the member on the left. Each *S* was given a booklet with 90 pairs grouped under 22 appropriate contexts. The *Ss* were given an explanation of free-association procedures as well as what was required of them in this task. They were told to work rapidly but carefully and were allowed to proceed at their own pace.

Imagery Rating Task

Method

Subjects. Forty males and 60 females from an introductory psychology class at the University of

volunteered to participate in this Each *S* received course credit for ng.

e. The left members of the pairs used sociative norming exercise were presented ... an imagery rating task similar to that of , Yuille, and Madigan (1968). Each word red beside a 7-point rating scale, with 7 ring the high end of the scale and 1 anchoring w end. The *S*s were given imagery rating ctions similar to those of Paivio et al. and allowed to work at their own pace.

ined Results

he two norming tasks yielded two values for each cue word, (*a*) a mean ion value between itself and its con- nd (*b*) a mean imagery rating. For h context, the cue word candidates were oss-classified on the association value and magery dimensions to obtain an item in each cell of the four-fold classification. The members of the classification were determined by selecting words from the extremes of the two dimensions. For the association value dimension, a word was classified as high if it had a mean of approximately three or better and low if its mean was two or less. On the imagery dimension, words with a mean of six or better constituted concrete words, while words with a mean of four or less constituted abstract words. Of the 22 contexts appearing in the norming exercise, 13 yielded candidates fulfilling these cross-classification requirements. Table 2 displays the means in a four-fold array for the 13 target triplets.

TABLE 2

CELL AND MARGINAL MEAN CLASSIFICATIONS OF TARGET ITEMS ON IMAGERY (*I*) AND ASSOCIATION VALUE (AV) DIMENSIONS

Association value	Imagery: Abstract	Imagery: Concrete	Mean AV
Low			
AV	1.68	1.39	1.53
I	2.73	6.23	
High			
AV	3.29	3.57	3.43
I	3.57	5.93	
Mean *I*	3.15	6.08	

MAIN EXPERIMENT

Method

Subjects. Thirty-two males and 32 females from an introductory psychology course at the University of Minnesota served as *S*s. Each *S* volunteered and received course credit for participating.

Design. A 2 × 2 × 2 randomized block design was used, with *S*s being blocked on sex. The other two between-*S*s variables were imagery (concrete and abstract) and association value (high and low). All three variables were crossed, and *S*s were assigned to the various combinations of imagery and association value by their order of appearance at the laboratory.

Materials. The list consisted of 20 triplets and contexts. The first 3 and last 4 groupings were filler items, while the middle 13 triplets were the target items resulting from the preliminary study. Each target triplet consisted of a context and two items that remained constant under all experimental conditions. The third member of each triplet varied according to the particular combinations of imagery and association value. For example, for the context WEDDING, the two constant items were DRESS and RING and the four cue words were as follows: concrete–high association value (AV), BRIDE; concrete–low AV, CANDLE; abstract–high AV, VOWS; and abstract–low AV, BEAUTY.

Procedure. The *S*s were run in groups of 4–10. They were read instructions regarding the nature of imagery and given an example of how they might form compound images for word triplets. In addition, they were told to use the provided context as a setting in which to form the image and to try to remember the triplet, such that when they were given one of the members of the triplet later, they would be able to recall the other two words. The *S*s were then given four examples for which they practiced forming images. The triplets were read to *S*s at a rate of one triplet every 15 sec., followed by a 1-min. subtraction task. After this, *S*s were read one word from each triplet and allowed 30 sec. to write the other two words.

Results

The first 3 and last 4 triplets on the list were fillers; therefore, only the remaining 13 triplets were included in the analysis. A 2 × 2 × 2 analysis of variance was performed on the recall scores, with the three between-*S* variables being imagery (abstract and concrete), association value (low and high), and sex (female and male).

The recall means for the four groups are shown in Table 3. The analysis of variance on these scores revealed two significant main effects: association value, F (1, 56) $= 58.43$, $p < .0001$; and sex, F (1, 56) $= 4.51$, $p < .04$. None of the other main effects or interactions was significant. The

TABLE 3

Mean Recall (*SD*s in Parentheses)

Association value	Imagery	
	Abstract	Concrete
Low	12.50 (5.06)	12.75 (5.50)
High	21.81 (2.86)	20.75 (4.65)

mean recall for the high-AV condition was 21.28 as compared to 12.62 for the low-AV condition, while the means for the two imagery conditions were quite similar (abstract = 17.16, concrete = 16.75). The mean recall for the females was 18.16 and for the males, 15.75. Interactions involving sex were not significant; therefore, this variable was not analyzed any further.

Discussion

The results of this experiment were quite unambiguous: Differences in recall were due solely to the association value of the cue word to the context, and the imagery value of the cue words had no effect. This finding would not be predicted from the conceptual-peg hypothesis, but it is entirely consistent with the constructive process view of imagery.

The important point to be realized from these results is that the critical factor for recall was not the concreteness of the cue words per se; rather, it was the function that the cue words served. In other words, in most of the studies demonstrating the importance of the concreteness of the stimulus term in cued recall (e.g., Paivio, 1965; Yarmey & Paivio, 1965), it is quite possible that the concreteness of the cue word was confounded with the term's capacity to elicit the context used to construct the image. In these studies, the concrete stimulus terms may also have been high in likelihood of eliciting the context used to construct the image, thereby yielding positive results. The present experiment did not confound these two factors and consequently demonstrated that the critical factor was the stimulus term's capacity to elicit the context.

Context is quite vaguely defined here and is meant to refer to the general setting from which the image can be reconstructed. When the statement "eliciting the context" is used, it is meant to refer to the act of restoring the setting used to construct the original image. Nothing is offered here as to how the process actually takes place; instead, the results demonstrate the importance of the reinstatement of the context.

The facilitatory effect of context or themes on the recall of prose is well known (cf. Bartlett, 1932; Bransford & Johnson, 1972; Dooling & Lachman, 1971). Generally, the position taken is that the context or theme is recalled along with a few details, and the remainder is reconstructed. The results of the present experiment indicate that the reinstatement of a context is also a critical element in recall through imagery, and they are consistent with the argument offered for the role of contexts in prose recall.

REFERENCES

Bartlett, F. C. *Remembering.* Cambridge, England: Cambridge University Press, 1932.

Bower, G. H. Mental imagery and associative learning. In L. W. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley, 1972.

Bransford, J. D., & Johnson, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 717–726.

Dooling, D. J., & Lachman, R. Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, 1971, **88**, 216–222.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Paivio, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. *Canadian Journal of Psychology*, 1963, **17**, 370–379.

Paivio, A. Abstractness, imagery and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 32–38.

Paivio, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.

Paivio, A., Yuille, J. C., & Madigan, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Pt. 2).

Yarmey, A. D., & Paivio, A. Further evidence on the effects of word abstractness and meaningfulness in paired-associate learning. *Psychonomic Science*, 1965, **2**, 307–308.

(Received August 27, 1973)

...rimental Psychology
No. 5, 845–849

LETTER SEARCH THROUGH WORDS AND NONWORDS BY ADULTS AND FOURTH-GRADE CHILDREN [1]

LESTER E. KRUEGER [2]
... College of the City University of New York

ROBERT H. KEEN
New York University

BELLA RUBLEVICH
City College of the City University of New York

Both fourth-grade children and adult Ss searched faster for target letters through common words than nonwords (scrambled collections of letters), through third-order pseudowords than nonwords. In fact, both fourth-graders and adults reduced their search time by nearly the same proportion: 10% through common words, 3% through third-order pseudowords. Better readers tended to be faster searchers, but there was no consistent relationship between reading ability and the ability to use the redundancy in words to speed letter research.

Adults can exploit their knowledge of word structure in a letter search task. Krueger's (1970b) undergraduate *S*s searched faster for target letters through real words and third-order pseudowords than through nonwords (scrambled collections of letters). In the present study, Experiment I examined whether fourth-grade children, whose knowledge and fam... ...ty with words is much more ..., yet has had some opportunity to ...velop, also would search faster through words than nonwords. Experiment I also examined whether fourth-grade children would follow the pattern shown by adult *S*s, by searching faster through repeated than nonrepeated letter strings (Krueger, 1970a) and searching faster through coherent sentences than the same words in scrambled lists (Krueger, 1970b).

Experiment I

Method

Subjects. Twenty City College of New York undergraduates, all paid volunteers, served as the adult *S*s. Ten pupils from Public School 192 in Manhattan served as the fourth-grade *S*s (median age: 9 yr., 7 mo.). English was the native language of all *S*s. All *S*s had 20/40 vision (corrected) or better.

Apparatus and materials. Stimuli were presented at 4 mL. intensity on a Gerbrands Model T-3B-1 tachistoscope. The exposure field was illuminated when *S* pressed a foot pedal at the beginning of a trial, and was extinguished immediately upon *S*'s response.

In Sets 1 and 2, each card contained a single (target) letter, centered above a column of five six-letter words or nonwords, typed in all-capitals, one to a line on a 4 × 6 in. white index card, in 10 pitch, IBM Courier 72. All typed lines were double-spaced.

In Set 1, the five six-letter items on each card were either all words or all nonwords. From the 140 common words used by Krueger (1970b, Experiment I), 14 words were chosen in a quasi-random manner so as to match as closely as possible the letter frequency distributions at each letter position in the 140-item set. Fourteen nonwords likewise were drawn from the corresponding 140-item nonword set. The 14 items chosen in each case served as repeated items; 120 of the 126 remaining items were selected for one-time use. A total of 192 cards were constructed: 72 with repeated words, 24 with nonrepeated words, 72 with repeated nonwords, and 24 with nonrepeated nonwords. The same item never appeared twice on a card. Each repeated item appeared on an average 25.7 cards. On 27 of the 72 repeated cards, and 9 of the 24 nonrepeated cards, the target letter at the top was not present in any of the five items below. When the target letter was present, it appeared equally often in either Line 1, 2, 3, 4, or 5, and always appeared in either Letter Position 2, 3, 4, or 5, but never 1 or 6. Each letter position was used equally often for words and nonwords and for repeated and nonrepeated displays. The 192 cards were divided into three equivalent subsets of 64 each, which were

[1] This study was supported by NIH Grant HD-04869-01-02-03 to the first author. The authors are grateful to Harry Rosenberg, Sheri Berenbaum, and Ora Ezrachi for assistance in tabulating the data, and to A. Rodriquez, principal, V. Lieber and H. Pierce, assistant principals, and C. Lazar and R. Stampp, teachers, at Public School 192, Manhattan, for their assistance in obtaining the fourth-grade *S*s.

[2] Requests for reprints should be sent to Lester E. Krueger, Department of Psychology, City College of the City University of New York, New York, New York 10031.

TABLE 1

EXPERIMENT I: RELATIVE REDUCTION IN SEARCH TIME FOR ADULTS AND FOURTH-GRADE *S*s

Group	Repeated vs. nonrepeated words and nonwords (Set 1)	Common words vs. corresponding nonwords[a] (Set 1)	Third-order pseudowords vs. corresponding nonwords (Set 2)	Coherent vs. scrambled passages (Set 3)
Adults	2.2%	8.4%	3.1%	1.0%
Fourth graders	8.5%	10.9%	2.5%	4.2%

[a] Based only on repeated-item displays.

randomized differently for each *S*; the order of the subsets also was rotated across the *S*s.

In Set 2, 16 cards contained five third-order pseudowords and 16 contained corresponding nonwords, selected from a set used by Krueger (1970b). No items were repeated. In each subset of 16 cards, the target letter was not present in 6 cards, and in the remaining 10 cards appeared twice in each of the five rows. The 32 cards were presented in a different random order for each *S*.

In Set 3, the card contained an extended passage, either in its original (coherent) version or in a randomly rearranged (scrambled) version (for further details, see Krueger, 1970b). The single letter was triple-spaced above the set of five lines, which were double-spaced, flush left. The 36 cards were presented in a different random order for each *S*; 18 contained coherent passages and 18 contained scrambled passages. In 6 of the 18, the target letter did not appear at all in the passage; in two cases each, it appeared in Lines 1, 2, and 4, and in three cases each, it appeared in Lines 3 and 5.

Procedure. The *S*s were told to search down the card for the letter shown at the top. Half of the *S*s were told to press the left hand button ("yes") upon finding the target letter, and the right hand button ("no") after searching the entire array without finding the target; for the other half of the *S*s, the hand assignment was reversed. The *S*s were told to search as quickly as possible, but to keep their errors to a minimum. A buzzer sounded when *S* pressed the incorrect button; all errors were excluded before search time was averaged.

Eight practice cards containing rare words or corresponding nonwords were presented first, followed by Set 1, then Set. 2 The *S*s next rated 56 items (a randomized list of 14 repeated words, 14 nonrepeated words, 14 repeated nonwords, and 14 nonrepeated nonwords) as to whether they had been presented more than once in Set 1. The adult *S*s next received the Nelson-Denny Reading Test, Form A, which measures vocabulary, reading comprehension, and reading rate. Eighteen of the 20 adult *S*s subsequently received Set 3, on either the same or a later date.

Each child received Set 3 at a later date, followed by the Gates-MacGinitie Reading Test, Survey D, Form 1M. Scores also were available on the Metropolitan Achievement Test, Elementary Level, Form H, administered at the child's school two to four months previously. Data from one child on Set 3 were excluded because of failure to follow the instructions properly.

Results

Search performance. The adults searched about twice as fast as the fourth-grade *S*s; to scan down an average 3.75 items in the Set 1 nonrepeated-nonword displays, for instance, took an average 3.38 sec. for adults and 6.75 sec. for fourth-grade *S*s. In Set 1, mean search time was significantly lower for adults than fourth graders, F (1, 28) = 49.62, $p < .001$, for word than nonword displays, F (1, 28) = 49.26, $p < .001$, and for repeated-item than nonrepeated-item displays, F (1, 28) = 18.30, $p < .001$. Significant interactions were found for Adults vs. Fourth-graders × Words vs. Nonwords, F (1, 28) = 5.11, $p < .05$, and for Adults vs. Fourth Graders × Repeated vs. Nonrepeated, F (1, 28) = 11.04, $p < .01$, reflecting the fact that search time decreased more for fourth graders than for adults when word or repeated-item displays were presented. When the decrease in search time is expressed not in absolute, but in relative terms, that is, as the percent reduction of baseline nonword or nonrepeated search time (see Table 1), fourth-graders still differed significantly from the adults on repeated vs. nonrepeated, t (28) = 2.86, $p < .01$, but not on words vs. nonwords.

In Set 1, the error rate was significantly lower for adults than fourth graders, 7.1% vs. 12.6%, F (1, 28) = 12.14, $p < .01$, and for word than nonword displays, 5.1% vs. 11.2%, F (1, 28) = 14.06, $p < .001$. Although the repeated vs. nonrepeated main effect (8.5% vs. 9.3%) was not significant, the Word vs. Nonword × Repeated vs. Nonrepeated interaction was significant, F (1, 28) = 6.21, $p < .025$, reflecting the fact that the repeated-nonrepeated difference was larger among nonword than word displays. No other interaction was significant.

The *S*s rated 56 words and nonwords from Set 1 as to whether they had occurred

once on the search displays. liott's (1964) table, d' values were d, based on the frequency with s answered "yes" to repeated items ite) and nonrepeated items (false rate). The mean d' value was significantly greater than zero, both for adults, 1.21, t (19) = 9.64, $p < .001$, and rth-grade *S*s, $d' = 1.00$, t (9) = 4.15, $< .005$; the difference between adults and rth graders was not significant. Mean was significantly greater for words than nwords, F (1, 28) = 14.79, $p < .001$, and word-nonword difference was virtually ie same, for adults, 1.51 vs. .91, and graders, 1.28 vs. .71.

In Set 2, the decrease in search time for third-order pseudowords (vs. corresponding nonwords) was significant for adults, t (19) = 2.29, $p < .05$, but not for fourth graders. The relative reduction in search time, shown in Table 1, was significantly smaller for the third-order pseudowords in Set 2 than the common words in Set 1, F (1, 28) = 20.10, $p < .001$. This factor did not interact significantly with adults vs. fourth graders. The mean error rate did not differ significantly between third- pseudowords and nonwords, for ther adults (3.1% vs. 4.7%) or fourth-grade *S*s (13.8% vs. 13.8%).

In Set 3, search was faster, but not significantly so, through coherent passages than scrambled ones, among both adults and fourth graders (see Table 1). Similarly, error rate was slightly, but not significantly lower for coherent passages, for adults (18.3% vs. 18.9%) and fourth graders (20.6% vs. 21.1%).

Correlations between search and reading measures. The various reading measures were highly and usually significantly ($p < .05$) intercorrelated. The correlations ranged from $r = +.47$ to $+.96$ among the children's measures (Metropolitan: word knowledge, reading; Gates-MacGinitie: speed, accuracy, vocabulary, comprehension), and $r = +.35$ to $+.89$ among the adult measures (Nelson-Denny: reading rate, vocabulary, comprehension). The high correlations indicate that there was a good spread in reading ability, both among the children and the adults. Thus, on the vocabulary test for the adults, the scores ranged from the 1st to the 93rd percentile, based on Grade 16 norms, with the median score at the 51st percentile. For the children, on the total reading ability measure of the Metropolitan Achievement Test, the scores ranged from the 18th to the 92nd percentile, based on Grade 4 norms, with the median score at the 48th percentile.

Search time invariably correlated negatively with the various reading measures, indicating that poor readers generally took longer to locate the target letter than did good readers. (The correlations were computed separately, and yielded quite similar results, for search time through repeated common words and search time through repeated nonwords.) Among fourth graders these correlations were significant ($p < .05$) for all four Gates-MacGinitie measures, ranging from $r = -.63$ to $-.73$, but not for either Metropolitan measure, ranging from $r = -.36$ to $-.56$. Among adults the correlations, ranging from $r = -.14$ to $-.32$, were not significant.

No trend was apparent at all, however, in the correlations between reading measures and the absolute or relative reduction in search time for various types of material (i.e., words, repeated items, coherent passages). The correlations were generally quite small, and were negative about as often as positive. The only significant correlations were between reading rate in adults and absolute ($r = -.55$, $p < .02$) and relative ($r = -.49$, $p < .05$) reduction in search time through coherent passages.

Discussion

As the significant d' scores indicate, both adult and fourth-grade *S*s discriminated rather well between the repeated and nonrepeated items. Schulman (1971) obtained similar results when he tested *S*s on their recognition memory of words they had scanned through for the letter A. Neisser (1964), on the other hand, reported that *S*s usually could not distinguish words that had appeared on a list from words that had not, when tested immediately after a scan. Neisser proposed a

hierarchical model in which nontarget letters are analyzed only so deeply that they may be discarded from further consideration. Neisser's well-practiced *S*s reported that they did not even "see" nontarget letters during search for a target letter. The present findings, based on the performance of relatively unpracticed *S*s who were set for accuracy as well as speed, indicate that nontarget letters not only are seen, but may even be organized and remembered as entire words or nonwords.

The adults searched about twice as fast as the fourth graders, which confirms previous work with third graders (Forsman, 1967) and fourth graders (Leslie & Calfee, 1971). The present finding that in the fourth grade the better readers search faster than do poor readers, though, contradicts two recent studies: Leslie and Calfee (1971), using second, fourth, and sixth graders, found no relationship between reading level and search rate for target words, and Katz and Wicklund (1972), using second and sixth graders, found no difference between good and poor readers in letter-search rate through nonsense items. On the other hand, Spring (1971), using third, fourth, and fifth graders, found longer latencies for poor than normal readers on the "same"-"different" judgment of two adjacent letters.

Since older *S*s and better readers exhibited superiority on the general skill of visual search, one would have expected them to exhibit an even greater superiority on the specific skill of exploiting their knowledge of English word structure to speed search for letters. As it turned out, however, reduction in search time through words was as great for children as for adults, as great for poor readers as for good readers. This result will be discussed more fully after Experiment II has been presented. Experiment II, in which 40 adult *S*s participated, provided another test of whether the exploitation of word structure in letter search is related to reading ability.

Experiment II

Method

Subjects. Forty City College of New York undergraduates served as paid *S*s.

Apparatus and materials. Search displays were presented in a light-tight wooden box containing a semitransparent (one-way) mirror (12 × 15 in.); *S* could see into the box whenever four 40-w. lamps inside were lit. On each trial, *E* first slid a plank containing the 8½ × 11 in. stimulus sheet into the box, and then closed a switch which illuminated the lamps and started a Standard Electric, Model S-1 timer (calibrated to .01 sec.). The *S*'s response extinguished the lamps and stopped the timer.

The sets of 140 common words and 140 corresponding nonwords used were the same as in Experiment 1. Each stimulus sheet contained either 25 six-letter words or 25 six-letter nonwords, typed in one column, single-spaced, in 12 pitch, IBM Prestige Elite 72. Sampling of items was random, except that no word or nonword appeared twice on a given list of 25. Each *S* saw 22 word and 22 nonword displays, which were randomly intermixed. The target letter was actually present in 15 of the 22 displays, appearing once each in Item Positions 2 and 4, and in all odd-numbered positions from 1 to 25 (for further details, see Krueger, 1970b).

Procedure. On each trial, after *S* saw the target letter and indicated he was ready, *E* turned on the timer and lamps. The *S* was told to press the right-hand ("yes") switch as soon as he found the target letter, or, if it was not present, to press the left-hand ("no") switch. The *S* received three practice trials. The *S* was informed when he pressed the incorrect switch, in which case his search time was discarded for both the word and nonword displays having a target at the particular location. After completing the letter search task, *S* received the Nelson-Denny Reading Test, Form A, and the Wellesley Spelling Scale, Form 1.

Results

Search performance. Mean search time, through an average 15.8 items, was 8.48 sec. for word displays and 7.71 sec. for nonword displays. The 9.1% relative reduction in search time for words (vs. nonwords), was significant, $t\ (39) = 4.8$, $p < .001$. The error rate was somewhat, but not significantly, higher for nonword (11.3%) than word (9.2%) displays.

Correlations between search and reading and spelling measures. As in Experiment I, the reading and spelling measures were highly and usually significantly ($p < .05$) intercorrelated (range: $r = +.26$ to $+.73$), indicating again a good spread in reading ability. Thus, on the vocabulary test, the scores ranged from the 2nd to the 99th percentile, based on Grade 16 norms, with the median score at the 72nd percentile. Also as in Experiment I, reading and spelling measures generally correlated negatively with search time for words and nonwords, $r = +.03$ to $-.34$. That is, better readers tended to search faster than poor readers. The only significant correlation, though, was that between reading rate and mean

se..h time through words, $r = -.34$, p...

...Experiment I, the correlations ...the reading measures and the ...ion in search time for words (vs. ...ords), were all positive, both for ...lute reduction ($r = +.03$ to $+.32$) and f.. relative reduction ($r = +.09$ to $+.39$). ..at is, better readers tended to take more ..dvantage of the familiar contexts of words ..n reducing search time. The only signifi...ant correlations, though, were those be...veen reading comprehension and relative ...duction in search time for words, $r = +$ ..39, $p < .02$, and reading comprehension ..d absolute reduction, $r = +.32$, $p < .05$.

General Discussion

Perhaps the most striking finding of the present study is that fourth-grade children made as much use of the redundancy in English and near-English letter sequences as did adults. Both adults and fourth-grade children speeded their search for letters by about 10% when nonwords were replaced by common words, by about 3% when nonwords were replaced by third-or..er pseudowords (see Table 1). These resul.. ..e in very nicely with the findings of ...nd Smith (1970), who found that the ..ility to use intraword redundancy in the identification of letters increased only up to the fourth grade, where adult-level skill was achieved.

The mastery of letter and word recognition skills by the fourth grade is very important, because the child is thereby freed to develop more complex skills in reading. Singer (1970) reported that the "system for attaining Speed of Reading undergoes a developmental shift from a predominance of visual perceptual abilities at the third grade to a more equitable organization of visual perceptual and word meaning factors at the sixth grade level [pp. 215–216]."

Of course, the visual perceptual component in reading may continue to improve to some extent well past the fourth grade, and better readers may ever surpass poor readers, as shown, for instance, by the fact that even among adults, the better readers tended to search faster for target letters. Among adults, however, reading ability accounted for no more than about 10% of the variation on search time. Among the fourth graders, by contrast, reading ability, as variously indexed, accounted for about 15–50% of the variation on search time.

References

Elliott, P. B. Appendix 1, Tables of d'. In J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964.

Forsman, R. Age differences in the effects of stimulus complexity and symmetrical form on choice reaction time and visual search performance. *Journal of Experimental Child Psychology,* 1967, **5**, 406–429.

Katz, L., & Wicklund, D. A. Letter scanning rate for good and poor readers in grades two and six. *Journal of Educational Psychology,* 1972, **63**, 363–367.

Krueger, L. E. Effect of frequency of display on speed of visual search. *Journal of Experimental Psychology,* 1970, **84**, 495–498. (a)

Krueger, L. E. Search time in a redundant visual display. *Journal of Experimental Psychology,* 1970, **83**, 391–399. (b)

Leslie, R., & Calfee, R. C. Visual search through word lists as a function of grade level, reading ability, and target repetition. *Perception & Psychophysics,* 1971, **10**, 169–171.

Lott, D., & Smith, F. Knowledge of intraword redundancy by beginning readers. *Psychonomic Science,* 1970, **19**, 343–344.

Neisser, U. Visual search. *Scientific American,* 1964, **210**(6), 94–102.

Schulman, A. I. Recognition memory for targets from a scanned word list. *British Journal of Psychology,* 1971, **62**, 335–346.

Singer, H. A developmental model of speed of reading in grades three through six. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading.* Newark, Del.: International Reading Assn., 1970.

Spring, C. Perceptual speed in poor readers. *Journal of Educational Psychology,* 1971, **62**, 492–500.

(Received August 31, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 850–862

THE LOCUS OF THE RETENTION DIFFERENCES ASSOCIATED WITH DEGREE OF HIERARCHICAL CONCEPTUAL STRUCTURE [1]

BENTON J. UNDERWOOD,[2] JOHN J. SHAUGHNESSY, AND JOEL ZIMMERMAN

Northwestern University

Constant-order paired-associate lists were used in which the numbers 1 were stimulus terms, and 24 nouns were response terms. The order of t nouns was varied across five lists to produce a different number of hierarchi conceptual levels in the lists. There were two degrees of original learning a three types of retention tests after 24 hr. The study–test method was us Learning rate was related directly to the degree of conceptual structure, retention was uninfluenced by structure. A further experiment showed th the direct relation between recall and structure found in an earlier study is t be attributed to the anticipation method in which information at recall is i an amount that is directly related to the conceptual structure.

In a previous study (Underwood & Zimmerman, 1973), 16 words were ordered serially so that a three-level conceptual hierarchical structure resulted. If *S*, in learning the list, followed the rules implied by this structure, placement of each word in its appropriate position within the list was possible. Other lists were constructed from the same words in a way such as to violate the appropriateness of successive conceptual levels. The purpose of this previous study was to determine the role of conceptual structure on the learning and retention of the lists. Two findings emerged. First, learning rate was related directly to degree of conceptual structure up to a point, and second, recall after 24 hr. was related directly to degree of structure. This latter finding conformed in general to the notion that associations learned in the laboratory which are compatible with already established associations will show less rapid forgetting than will be the case for associations which are in conflict with established habits. In the previous study, this latter case was represented at the extreme by the 16 words presented in random order. It was presum that for such a list the long-established onceptual habits would interfere with th propriate ordering of the words at re In fact, however, the overt-error da gave no evidence that the poorer recall the random list than of the structural list was due to such interference. Thus, the reasons for the differences in recall remained obscure.

The intent of the present study was identify more precisely the characteristic of the memory for structured and unstructured lists. By so doing it was believed that the characteristic or characteristics responsible for the differences in retention might be isolated. Two different levels of learning were used as a means of varying the degree to which the conceptual structure had become a part of the memory for the lists. Since the utilization of a conceptual structure to mediate item placement at the time of retention may be critically time dependent, three different types of retention tests were used, namely, paced and unpaced recall, and unpaced associative matching.

Main Experiment

Method

Lists. Five different lists were constructed, all from the following 24 words: *robin, owl, bobolink, trout, guppy, sturgeon, apple, lemon, fig, rose, lilac, marigold, beer, rum, sherry, milk, soda, cocoa, diamor*

[1] This research was supported by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract N00014-67-A-0356-0010, Contract Authority Identification No. NR 154-321.

[2] Requests for reprints should be sent to Benton J. Underwood, Department of Psychology, Northwestern University, Evanston, Illinois 60201.

opal, sapphire, iron, brass, tungsten. These 24 words, presented as in the order listed above, formed the most highly structured list, to be called List 5. It will be noted that at the lowest conceptual level in List 5 there are three instances of each of eight concepts, birds, fish, fruit, flowers, alcoholic beverages, nonalcoholic beverages, precious stones, and metals. At the next conceptual level there are four concepts, animals, plants, beverages, and minerals. At the third level, living and nonliving things divide the list in half.

In List 5 the three instances within each concept were ordered serially (as above) such that the first instances had high frequency in the Battig–Montague (1969) norms, the second medium frequency, and the third, low frequency. Given this ordering the *S* could, in a manner of speaking, run off the three instances within a concept according to a frequency rule. In the previous study (Underwood & Zimmerman, 1973), this variable, with two instances of each concept, did not influence learning. However, there were reasons to believe that the earlier lists were not entirely satisfactory for a test of the frequency rule and so this variable was included again in the present experiment.

The nature of the other four lists may now be described. For List 4 the conceptual structure remained the same as in List 5, but the order of the three instances within each of the eight concepts was randomized to neutralize any influence of a word-frequency rule. In List 3 the order of the animal names was arranged so that the concepts birds and fishes were not appropriate for three successive items, but the concept animal was appropriate for a block of six words (*robin, trout, sturgeon, owl, guppy, bobolink*). The same was true for plants, beverages, and minerals. Therefore, the four intact concepts in List 3 could mediate placement only within a block of six positions. For List 2 only the living–nonliving distinction was maintained so that the implementation of this distinction by the *S* would restrict placement to halves of the list. Finally, in List 1, the ordering was random so that no conceptual mediation of placement of groups of words was possible.

The lists were presented as constant-order paired-associate lists with the numbers 1–24 in order as stimuli and with the words as response terms. In the previous experiment, this procedure gave results which were essentially equivalent to those found when the words were presented as a true serial list.

Conditions. One variable, of course, was defined by the five lists as explained above. A second variable was the degree of learning prior to the 24-hr. retention interval. Half of the *S*s learned to a criterion of 12 correct responses on a single trial, half to a criterion of 20 correct responses on a single trial. A third variable was the nature of the retention test. Half of the *S*s were given a paced-recall test, the rate being the same as used during learning. For these *S*s, recall was followed by relearning to one perfect recitation but with a minimum of three relearning trials after the recall trial. Half of the *S*s were given unpaced retention tests consisting of two steps. Initially, a sheet was provided the *S* on which the numbers 1 through 24 were listed with a blank after each. The *S* was given unlimited time to write down all of the appropriate response terms he could, guessing being encouraged. Following this step, a list of the 24 response words was provided the *S* on a second sheet and he was asked to match each word with a number, using each only once. He was required to fill each blank with a word even if it involved guessing.

The five lists, two degrees of learning, and two types of retention tests resulted in 20 different conditions. Four further conditions were added to provide controls for possible differential performance on the postcriterial trials of learning. The study–test method was used during learning, and although the use of this method is normally expected to minimize differences on postcriterial trials as a function of rate of learning, it was believed necessary to provide some minimal information about the matter. Consequently, four groups were given immediate retention tests after achieving the criterion of 20 correct responses on a single trial. Two of these groups learned List 1, and two learned List 5. For each list, one group was given paced recall and relearning, one unpaced recall, followed by matching.

There were 24 groups of college-student *S*s, one corresponding to each of the 24 unique conditions, with 18 *S*s in each group. Assignment to a particular group was made from a schedule containing 18 blocks of conditions, with each condition occurring once within each block. A different random order was used for the conditions within each block.

Procedure. All lists were presented for alternate study and test trials at a 1.5-sec. rate for both. When the appropriate criterion was achieved, the *S* either was dismissed from the laboratory (to return 24 hr. later) or was given an immediate retention test. On the immediate paced tests the experimenter stopped the memory drum, told the *S* that he would now have another test trial, with further study and test trials to follow, and that he should try to get as many correct as possible. For the unpaced tests (both immediate and delayed) the recall sheet was given the *S* and he was asked to follow the printed instructions on the sheet as the experimenter read these instructions aloud. He was given a second sheet (described earlier) for the matching test. All *S*s having 24-hr. tests were reminded that they had learned a list of words the previous day and that a test of their memories for this list was the purpose of the present session.

Finally, all *S*s having the 24-hr. retention tests were given an open-ended questionnaire concerning rehearsal activities they might have engaged in over the 24 hr. They were asked to describe any experiences they had with the words during the 24-hr. period, whether they had rehearsed or thought about the words, and so on. Honesty and accuracy were emphasized since (the *S* was told) such replies were valuable in helping to understand the nature of memory.

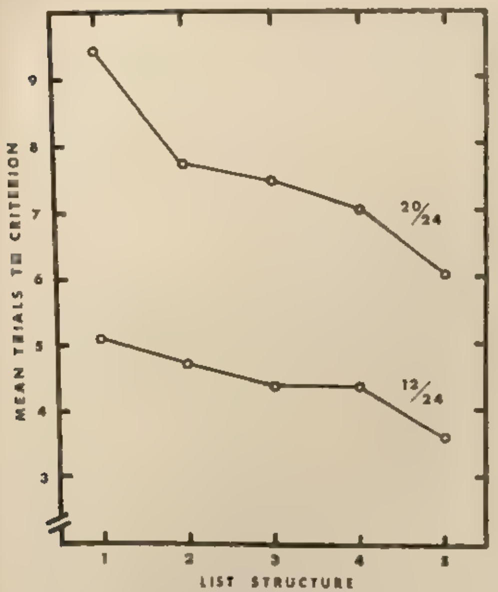


FIGURE 1. Mean number of trials to reach two different criteria of learning (12 correct on a single trial, and 20 correct), for lists varying in conceptual structure from low (List 1) to high (List 5).

Results

Original learning. For an examination of learning as a function of list structure, the *S*s who would subsequently receive different retention tests were combined to provide 36 *S*s having had each of the five lists at each criterion of learning. The data are plotted in Figure 1 in terms of the mean number of trials required to reach the two criteria. For convenience, the five lists are equally spaced along the baseline, indicating increasing structure from List 1 to List 5. It can be seen that as list structure increased, trials to learn decreased. This is true for both criteria, and summed across the two, the effect is significant statistically, F (4, 350) = 8.08, $p < .01$. Although the influence of list structure appeared to be somewhat greater for the higher criterion of learning than for the lower criterion, the interaction was not reliable, F (4, 350) = 2.14, $p > .05$. In the previous study (Underwood & Zimmerman, 1973), learning rate increased from List 1 through List 3 with no further increase for Lists 4 and 5. It is not known whether this difference is due methods of learning (anticipation vers dy-test, or to list differences. The was some evidence in the earlier study some of the instances defining the low level concepts were not always unders by the *S*.

It will be remembered that 5 differed from List 4 only in terms of ordering of the three words within ea the eight concepts. For List 5 th rds were ordered from high to low in ns of the frequency with which the i ces were produced to the concept nam the Battig and Montague (1969) normati tudy. In List 4, the three words wer domized. As seen in Figure 1, List 5 was ned more rapidly than was List 4, and was true for both criteria. An analysis variance with Lists 4 and 5 as one varial , and the two criteria as the other, showe he difference between the lists to reliable, F (1, 140) = 5.45, $p < .05$.

The effect of word frequency r all lists was also examined. Three score were obtained for each *S*, these repre nting the number of times the eight hig frequency instances were given correctly, the number of times the eight medium-frequency instances were given correctly, and the number of times the eight low-frequency instances were given correctly. Across all lists combined for the lower criterion of learning the three means were 13.99, 11.02, and 9.84, for the high-, medium-, and low-frequency instances, respectively (F = 87.59). These differences essentially disappeared at the higher criterion of learning for Lists 4 and 5, but were still quite evident for the other three lists. It appears, therefore, that one of the characteristics of initial learning of these lists by the study–test method was that either because of ease of learning of words with high frequencies, or because of priority effects, or both, the *S* acquired the high-frequency instances initially regardless of the conceptual structure of the list.

Still another finding could be interpreted as a word-frequency effect. Some observations of the experimenters suggested that the learning of the concepts differed fo men and women. The learning of List

d 5 was evaluated to see if these ob- ns had validity. The 144 *S*s (both of learning) consisted of 77 men 7 women. The number of correct re- nses within each of the eight concepts determined for each group. These are tted in Figure 2 for two reasons. First, hey show the nature of the serial-position urve which obtains even for the highly structured lists. Second, they indicate differential learning of certain concepts by men and women. Overall, the learning of the two groups did not differ ($F < 1$), but the interaction between concepts and sex was reliable, F (7, 994) = 3.13, $p < .01$. Three concepts produced substantial differences in learning between men and women, namely, *flowers*, *alcoholic beverages*, and *precious stones*. It may be that cultural experiences have resulted in women having a greater familiarity for words representing *flowers* and *precious stones* than is true for men, and that the opposite is true for *alcoholic beverages*. In any event, these results are similar to those reported by Bousfield and Cohen (1956) when they used free-recall learning of groups of conceptually related words.

Figure 2 suggests that the words which represented the first occurrence of a nonliving concept (alcoholic beverages) caused a distinct break in the position-performance curve. However, this is primarily a function of the particular words involved. For Lists 2 and 3, where the living–nonliving break also came between Positions 12 and 13, there was no evidence of discontinuity between the two positions.

Overt errors. The first overt error analysis to be presented dealt only with the number of errors across lists. The 360 *S*s from whom the acquisition data were derived (as plotted in Figure 1) were used. The error measure was the proportion of times an error was produced per opportunity (number of overt errors/number of omissions plus overt errors). For the 10 groups this ratio varied between .10 and .19. The values were greater statistically for the high-criterion groups than for the low-criterion groups, F (1, 350) = 8.33, $p < .01$, the means being .16 and .12. The errors

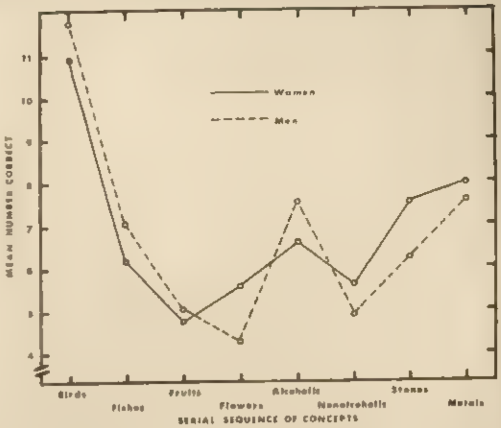


FIGURE 2. Learning of the highly structured lists as a function of the particular concepts (and their positions in the list) by men and women.

increased with list structure (.11, .11, .15, .15, .16), F (4, 350) = 4.08, $p < .01$. The F for the interaction was less than one. Undoubtedly, some correct responses resulted from guesses within concepts, particularly for the more structured lists, but in view of the relatively low proportion of overt errors to cases of not responding, it does not seem likely that mere guessing is heavily involved in the correct-response data.

For Lists 4 and 5, overt errors can be identified as having one of four levels of appropriateness in terms of the conceptual structure for these lists. These will be described and illustrated. An error may occur within the appropriate block of three instances representing the same concept. Saying "guppy" to one of the two stimuli to which it was an incorrect response in the block of three fishes was such an error, and will be called here an error at Level 1. Six words were involved in the animal block; if "guppy" was given to one of the three stimulus terms appropriate for birds, it was defined as a Level-2 error. Twelve words were included in the living block; if "guppy" was given to any of the six stimuli paired with plants, it was called a Level-3 error. Finally, if "guppy" was given as a response to any of the 12 stimuli paired with nonliving objects, it was called a Level-4 error. All errors were classified

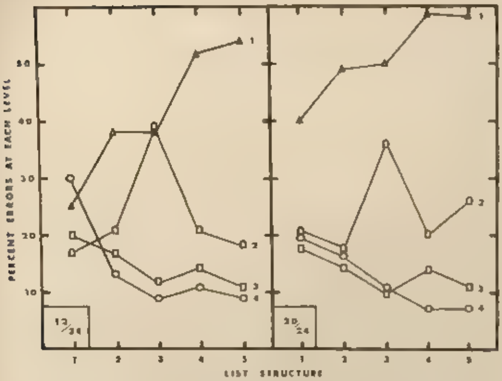


FIGURE 3. Locus of overt errors as a function of list structure and degree of learning. (The numbers 1, 2, 3, and 4 represent an increase in the discrepancy between the locus of an error and its correct position as defined by the conceptual structure of Lists 4 and 5.)

into one of these mutually exclusive categories for Lists 4 and 5. Conceptual level and position within Lists 4 and 5 are perfectly confounded. Position is also tied to stimulus number. Therefore, any analysis of the appropriateness of overt errors to the conceptual structure must consider the number of such errors which were produced by position per se. To handle this problem, the overt errors made in learning all five lists were categorized in exactly the same way as was done for Lists 4 and 5. For example, if, for List 1 (no structure), a misplaced response occurred among the first three pairs, it was classified as a Level-1 error, just as was done for List 5. Or, if an error made to the stimulus-term 9 was an appropriate response for the stimulus-term 3, it was classed as a Level-3 error. To summarize: for all lists the errors were classified as falling within one of eight blocks of 3 positions (Level 1), one of four blocks of 6 positions (Level 2), one of two blocks of 12 positions (Level 3), and, all others (Level 4).

The errors made by each *S* were allocated to the appropriate level and the percentage of errors in each level was determined. The means of these values were used to construct Figure 3. The sums of the percentages for a given list do not always equal 100 since a few *S*s produced no overt errors, and the means shown in Figure 3 were always based on $N = 18$. It should be noted that list structure is along the abscissa, with the error levels (1, 2, 3, 4) as the plotting parameter. The left panel represents the data for the *S*s who learned to a criterion of 12 correct, the right panel for those who learned to a criterion of 20 correct. Several features may be noted. For Lists 4 and 5, the greatest percentage of errors by far was that for errors falling into Level 1. Over half the errors made by the *S*s learning these lists represented errors that were approrpiate to the narrowest conceptual categories. The Level 1 values for List 1, values which should represent in relatively pure form the effects of position as such, are at 25% for the lower criterion, 40% for the higher criterion. In List 3 the narrowest conceptual category consisted of six words. If conceptual structure were used by the *S* in the placement of items in this list, he could reduce his possibilities to one of six positions. As may be seen in Figure 3, the maximum number of Level-2 errors occurred in learning List 3, these errors reflecting placement within a block of six words. For List 1, the random list, the maximum frequency of errors for the lower criterion was in Level 4, which means giving a response in one half of the list which actually belonged in the other half. There is only slight evidence that the living–nonliving distinction influenced the errors, this being shown in the fact that fewer errors at Level 4 were made by the *S*s learning List 2 than by those learning List 1. As would be expected, the major difference in the errors for the two criteria of learning was that there was an increase in Level-1 errors from the lower to the higher criterion, with the increase being the greatest for the lists with low structure. In summary, Figure 1 showed that learning was related directly to the number of appropriate conceptual levels involved in the list; the error data of Figure 3 indicate that these conceptual levels aided learning because they limited the number of possible numerical stimuli or positions for which a particular word was appropriate.

There is the possibility that with zero conceptual structure (List 1), the presence of conceptually related words actually interfered with learning. If such interference was present, it was not manifest in the error data. The errors made in learning List 1 were divided into two categories, those which were given to a stimulus for which another instance of the same concept was appropriate, and those which were not appropriate in the above sense. The concepts used were the eight with the three instances each. To illustrate: saying "guppy" to the stimulus appropriate for *trout* would be viewed as evidence for interference; giving "guppy" to the stimulus for *milk*, would not. Errors which would constitute evidence for interference constituted 7.6% of the total errors. Chance responding would be expected to yield 8.7% of such errors. This indicates that interference resulting from the conceptual relationships among the words was minimal in the unstructured list

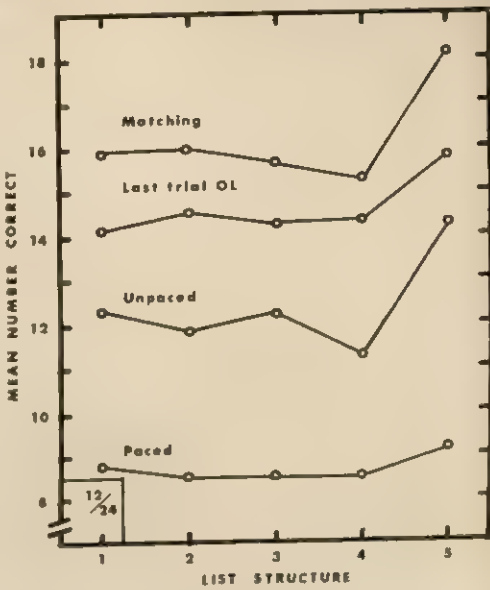


FIGURE 4. Number of correct responses on the various retention tests after 24 hr. following original learning to 12 correct responses. (Number correct on the last learning trial is also shown.)

Retention. The retention data will be presented first for the *S*s who learned to a criterion of 12 correct on a single trial. The results for all three retention tests are shown in Figure 4. To replicate the earlier study (Underwood & Zimmerman, 1973), performance should have increased as list structure increased. It is obvious from Figure 4 that list structure had little effect on retention. The means for the paced 24-hr. recall showed little variance ($F < 1$). For unpaced recall, performance was about the same for Lists 1 through 4, with an increase for List 5. However, the F for the recall for the five lists was only 1.37. It can be seen that performance on the last trial of learning was higher for List 5 than for the other four lists. It is likely, therefore, that the retention of List 5 for both paced recall and matching reflected this higher degree of learning, although it is not clear why paced recall for this list should not have been more influenced by this factor than it was. The fact that matching exceeded unpaced recall indicates that the *S*s were unable to recall some response words but could pair them appropriately when given the matching test. Again, however, the lists did not produce differential performance on the matching task ($F = 1.46$).

No immediate retention test was given to *S*s learning to the lower criterion so it is difficult to calculate the amount of forgetting which took place over the 24 hr. However, the number correct on the last trial of original learning does give some basis for a rough estimate. An immediate test would probably have shown some loss due to the fall normally found after reaching a criterion. As will be shown later, it would not be unreasonable to expect this loss to be at least one to two items. If so, the amount of forgetting was between 30% to 40% for the paced recall for the five lists. The critical fact is that list structure was unrelated to the amount lost over 24 hr

The next data to be examined are the retention scores for those *S*s who learned to a criterion of 20 correct responses. These are shown in Figure 5. For these comparisons, immediate tests are available for Lists 1 and 5 (dotted lines). It may be noted that the mean number correct on the last test trial of original learning varied between

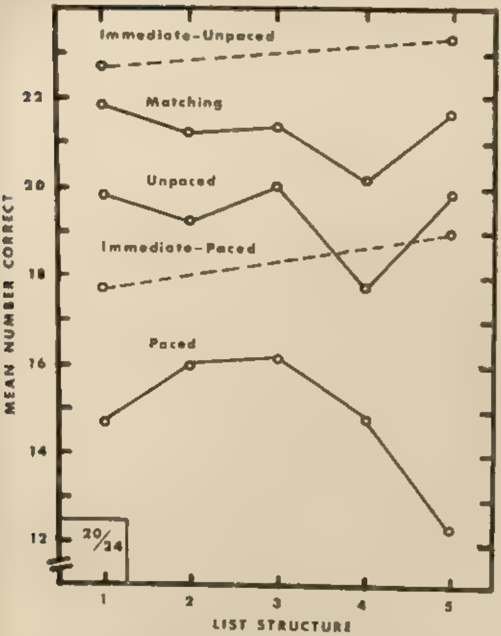


FIGURE 5. Number correct on the various 24-hr. retention tests following original learning to 20 correct responses. (Number correct on the immediate tests for Lists 1 and 5 is also shown.)

21 and 22 correct responses. Paced recall means varied between 12.28 for List 5, and 16.11 for List 3. Yet, overall the F was only 1.74. Because of the apparent difference in paced recall as a function of list structure for the two levels of original learning (Figure 4 versus Figure 5), an analysis was performed on these 10 groups. List structure was not a reliable source of variance ($F < 1$), level of original learning was ($F = 93.14$), but the interaction was not ($F = 1.66$). Thus, in spite of apparent differences when viewed graphically, both levels of original learning resulted in the same conclusion, namely, that list structure and paced recall were not related. Figure 5 suggests the possibility that there was an interaction between immediate paced recall and 24 hr. paced recall for Lists 1 and 5. A statistical analysis fails to confirm the reliability of this interaction, F (1, 68) = 3.34, $p > .05$. Using the immediate paced recall as a base, forgetting for paced recall is estimated at 25% across the lists as a whole.

It is obvious that forgetting als curred over the 24 hr. for the *S*s give paced recall. Again, however, list st re was not related to the amount of r under the unpaced procedure ($F < 1$ d the matching scores almost perfect flected the unpaced recall except the l was a little higher. In summary: th ention data for both levels of learning, paced, unpaced, and matching, all lead e conclusion that there was no syste relationship between list structure an rning using the study–test method.

The number of misplaced resp ses at recall was examined for all pac -recall groups. These numbers did not d er as a function of the criterion of learn g, but for both levels of learning the nu ber of misplaced responses at recall incr sed as list structure increased, F (4, 170) = 5.81, $p < .01$. This same relationship w found during learning and so is not a phenomenon peculiar to recall. For unpaced recall, however, list structure did not i luence the number of incorrectly-paired responses given.

As noted earlier, during learning of the constant-order paired-associate lists, very clear serial-position effects were present. These position curves were in evidence for all of the types of retention tests. Furthermore, there was a strong relationship between the number of times an item was given correctly in original learning and the number of times given correctly at recall, which is to indicate that the position effects remained essentially constant from learning to recall. For example, for the lower criterion, the correlations between number of times correct in original learning and number of times correct at recall varied between .77 and .91 for the five lists. Degree of original learning was obviously a powerful determiner of the items which were recalled.

Relearning. Differences in relearning as a function of list structure appeared on the first test trial after recall, and in terms of trials to reach a criterion of one perfect recitation, were reliable, F (4, 170) = 4.14, $p < .01$, with, of course, rate of relearning being directly related to structure. Thus,

list structure, which did not influence recall, q[illegible] reinstated during relearning the in[illegible]ce it had had during original learning.

[illegible]earsal and recall. The *S*s were given an open-ended questionnaire concerning th[illegible] rehearsal activities over the 24 hr. The replies to these questionnaires were r[illegible]ed on a 9-point scale for the amount of [illegible]earsal implied in the protocols. These [illegible]ngs were carried out independently by [illegible]e different people. Interrater reli[illegible]ities were determined for 18 subblocks [illegible]20 *S*s each. Of the 54 correlations possible, 11 were between .71–.80, 29 be[illegible]een .81–.90, and 14 were .91 or greater.

There were 20 groups of *S*s of 18 each having 24-hr. recall. All correlations between rated rehearsal activity and retention were positive, varying between .14 and .81. There was no relationship between rated rehearsal activity and list structure, and the magnitude of the correlations between rehearsal and retention did not vary systematically as a function of list structure.

The positive relationship between reported rehearsal and retention allows several alternative interpretations. For example, rehearsal may have increased retention, or, as another interpretation, *S*s with good retention may have rehearsed. The concern of the present study was whether or not rehearsal differed as a function of list structure. Since it didn't, it seems unlikely that the failure to find an influence of list structure on retention could be due to differential rehearsal. Also, if the present evidence on rehearsal can be generalized to the earlier study (Underwood & Zimmerman, 1973), it seems unlikely that the positive relationship between list structure and retention as reported for that study was due to differential rehearsal.

Discussion

In the previous study, paced recall and list structure were directly related and the forgetting over 24 hr. for the list with the highest structure was only 5%. In the present study, list structure was unrelated to either paced or unpaced recall, and the forgetting for the highest criterion of learning used was estimated at 25%. The discrepancy between these two studies is the topic of this preliminary discussion.

The contradiction in the retention results for the two studies pointed immediately to a method difference as the likely source for the contradiction. In the previous study the anticipation method was used. During the recall trial, therefore, the *S* would be informed of the particular concept whose instances were appropriate at the moment for the structured lists. If, for example, he remembered there were six animal names in succession, the appearance of the first instance would inform him that the next several positions also contained animal names. With the study–test method, the *S* would not be given this information. If he did not give the first instance of a concept at the appropriate point (to the correct stimulus), his following responses would be incorrect unless he remembered a particular response term associated with a stimulus term within the series of concept instances. There were, in fact, some cases in which *S* did give three to six correct response words in the correct order but which were scored as wrong because the initial response in the series was not paired with the appropriate stimulus. This did not occur frequently (and the data show that scoring these as correct did not change the basic conclusion), but the uncertainty felt by the *S* may have prevented him from responding overtly. List structure clearly influenced learning and relearning, but some portion of the learning which allowed the *S* to align response terms and stimulus terms correctly during learning must have been forgotten over the 24 hr. It is possible that the numerical stimulus terms were not always used as the effective stimulus terms and that the lists were treated more as serial lists than paired-associate lists, although what this means theoretically is not known. It appears that if *S* had learned, for example (for Lists 4 and 5), that number 1 was paired with the first bird instance, number 4 with the first of three fish instances, and so on, that this part of the memory was lost over 24 hr. Otherwise, there is no reason why providing this information at recall (as is done in part under the anticipation method) should result in better performance.

It seemed necessary to make a test of the notion that the discrepancy in the recall results for the two experiments was due to a difference in methods. In the auxiliary experiment, to be reported now, Lists 1 and 5 were presented for anticipation learning, with re-

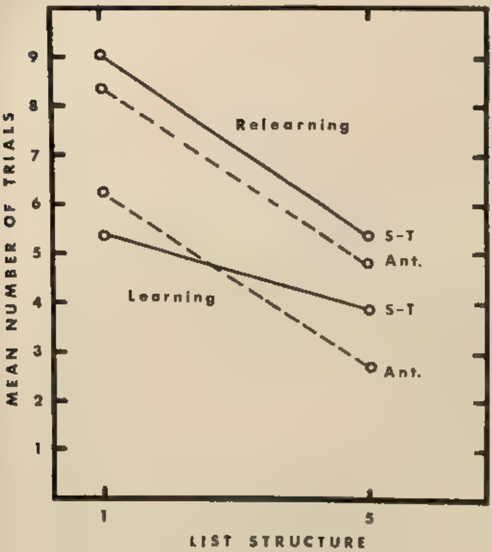


FIGURE 6. Learning and relearning scores for Lists 1 and 5 as a function of the study-test (S-T) and anticipation (Ant.) methods.

call taken after 24 hr. The expectation was that recall of these two lists would differ, with List 5 giving better recall than List 1.

AUXILIARY EXPERIMENT

Method

Lists 1 and 5 were presented at a 1.5:1.5-sec. rate for anticipation learning. The criterion of original learning was 12 correct responses on a single trial. Paced recall occurred after 24 hr. with relearning carried to one perfect trial, but a minimum of three trials beyond the recall trial. Each list was learned by a separate group of 18 *S*s assigned to one of the two lists by a block-randomized schedule.

Results

In presenting the results, comparisons will be made with the two groups of *S*s from the major experiment having the study–test method and paced recall, and who had learned Lists 1 and 5 to a criterion of 12 correct responses before the retention interval. There were 18 *S*s in each of these two groups. Since the auxiliary experiment was conducted after the major experiment, it is not known if the groups (study–test vs. anticipation) represent the same or equivalent populations. Therefore, differences in the levels of performance will not be readily interpretable, although interactions between the methods and lists should be meaningful.

The mean numbers of trials to learn to a criterion of 12 correct, and the mean numbers of trials to relearn to one perfect trial, are shown for the four groups in Figure 6. For original learning, list structure appears to have a greater influence for anticipation learning than it does for the study–test method. List structure is significant ($F =$ 25.33), as is the interaction, F (1, 68) = 4.19, $p < .05$. The interaction between list structure and method does not occur during relearning, indicating that the method's influence on learning is confined to the early stages.

It will be remembered that for the study–test method the frequency of concept instances was directly related to learning. This was quite evident at the lower criterion of learning. For List 5, under the anticipation method of the auxiliary experiment, the reverse was found. More specifically, two findings held across all eight concepts of three instances each. First, the initial word of the three (the high-frequency instance) was never given correctly more times than the second instance (the medium-frequency instance). Second, the third word in each of the eight concept triads (the low-frequency instance) was always given correctly more times than the first instance of the succeeding concept. For List 1 under anticipation learning, the effect of word frequency was in evidence just as was true for all of the lists learned by the study–test method. The above facts would indicate that item learning under the two methods would be more reliable for List 1 than for List 5. The product-moment correlations for item learning were .83 for List 1, and .58 for List 5. The positive relationship for List 5 reflects the commonality in learning by the two methods produced by serial position of the words, differences which were apparent under both methods.

Mean overt errors per opportunity were greater in learning under the anticipation method (.22) than under the study–test method (.13), and the difference was re-

lio' . , F (1, 68) = 9.24, $p < .01$. This may rep . . at a greater tendency to guess under th . . ticipation method than under the st . test method. However, of the overt er . . made under the two methods, the pe . entage of these errors within the ap- pr . riate concept position (Level-1 errors as . escribed earlier) was about the same for L . . 5, being 53% for anticipation and 54% f. the study–test method.

. he mean numbers of correct responses . ecall are shown in Figure 7, along with . mean number correct on the last learn- . trial. Under the anticipation method, . call was directly related to list structure, . (1, 34) = 4.97, $p < .05$. The inter- . tion between lists and methods for recall was also reliable, F (1, 68) = 4.44, $p < .05$. Recall for List 5 under the anticipation method was higher than performance on the last learning trial, 24 hr. earlier. To some extent, these comparisons are all in error, a matter which needs discussion.

Consider first the recall of List 1 under the two methods. A conclusion from an inspection of Figure 7 might be that recall . superior following anticipation learning . that following learning by the study– . st method. However, two factors must . considered. First, the criterion fall . ich may occur under the study–test method, and second, the learning which occurred on the last trial under the antici- pation method. This last issue can be handled directly. A multiple-entry prob- ability analysis (Underwood, 1964) was carried out to project performance to the hypothetical next trial under the anticipa- tion method. The mean expected value was 14.89. Since recall was 11.67 items, forgetting over 24 hr. was 3.22 items, or 22%. To incorporate the criterion fall into the calculations for List 1 for the study– test method requires a rough estimation of values. In the main experiment, groups were given a recall test immediately after reaching a criterion of 20 correct responses. On this immediate test under paced recall, performance fell from a mean correct of 20.89 correct on the last test trial to 17.72 correct on an immediately following test trial. This represents a loss of 3.17 items

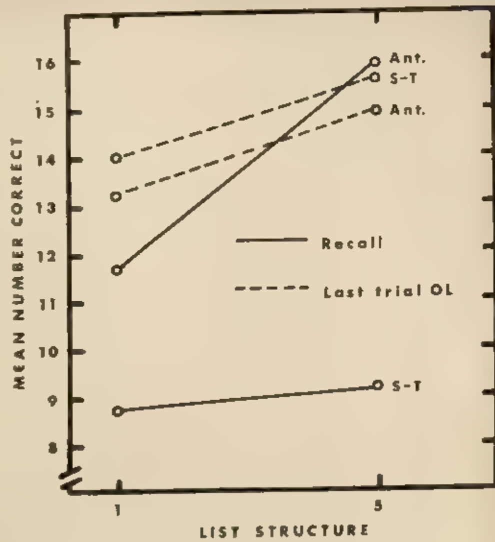


FIGURE 7. Number correct on the last trial of original learning and number recalled after 24 hr. for the study–test and anticipation methods.

which is referred to here as the criterion fall. None of the groups had an immediate test following learning to a criterion of 12 correct responses. Assume, however, that the group having the study–test method and learning to a criterion of 12 correct responses actually forgot 22%, the same as forgotten by the group having the anticipation method. This would require a criterion fall of 2.75 items. In light of the criterion fall of 3.17 items shown by the group learning to 20 correct responses, a fall of 2.75 items for those learning to 12 correct would not seem to be seriously in error. Given this assumption, the con- clusion is that for List 1, the amount of forgetting shown under the two methods is roughly the same.

Turning next to List 5, it must first be noted that it was not possible to project "next trial" performance for the Ss learning under the anticipation method because several of them learned in one or two trials. As an estimate, however, the percentage projection obtained for List 1 was used. This gives a mean expected next-trial value of 17.03 items. Since recall was 15.69 items on the average, the loss over 24 hr. was about 7%. Deriving an expected loss

TABLE 1

Comparison of Study–Test and Anticipation Methods with Regard to Concept Instance Learning and Recall for List 5

Correct responses	Anticipation		Study–Test	
	Last OL	Recall	Last OL	Recall
Three	117	162	210	102
Two				
1 & 2	22	14	18	18
1 & 3	28	20	18	10
2 & 3	62	68	18	8
One				
1	6	2	11	16
2	13	13	2	1
3	21	12	5	9

Note. See text for complete explanation.

for List 5 following study–test learning (using the value of 2.75 for the criterion fall) gives a loss of 3.75 items from 12.92, or 29% forgetting. These values must be considered approximations, but they lead to the conclusion that the loss was about the same under the two methods for the unstructured lists, but that the anticipation method led to better recall than did the study–test method for the highly structured list. Both the estimate of loss for List 1 for the anticipation method (22%) and the loss for List 5 under the same method (7%) correspond to the findings of the earlier published experiment using this method (Underwood & Zimmerman, 1973).

The sources of the differences in the retention for List 5 following the learning by the two methods needs more detailed examination. The data which seem to aid in reaching decisions about the sources of differences are shown in Table 1. The data sheets were examined for the last trial of original learning and for the recall trial. A tabulation was made of the number of correct responses which resulted from producing all three instances of a concept, the number which resulted from producing only two instances of a concept correctly, and the number which resulted from giving only one correct response from among the three possible for each concept. When two correct responses were given, they were broken down by position within the three possible positions (1 & 2, 1 & 3, 2 & 3). When a single correct response was given, the number falling in each of the three positions was noted. All of this information is given in Table 1 for the last original learning trial, and for the recall trial, under both methods. For example, on the last learning trial under the anticipation method, there were 39 cases in which all three instances were given correctly, resulting in 117 correct responses, the latter value being shown in Table 1. For this same condition there were 13 cases in which a single correct response was given for a concept and which consisted of the second of the three instances in the series of three.

Several facts are to be noted in Table 1. First, under the anticipation method there was an actual increase over 24 hours in the number of cases in which all three instances were correctly given (117 to 162). In view of the learning which may have taken place on the last anticipation trial of learning, this increase must be viewed cautiously. Nevertheless, it is in marked contrast to the results found with the study–test method, where there was a loss of 108 (210 − 102) complete triads. The second fact to note is the difference between the two methods in both learning and recall when less than three instances were given correctly. Under the anticipation method, the most probable positions for correct responses when two were given are Positions 2 and 3, and Position 3 when a single correct response was produced. Comparable relationships do not exist for the study–test method, where there was more or less constancy among the positions. The third fact is that if the number of cases in which the first instance (by position) of a concept was given correctly was calculated (regardless of the outcome for the two following instances) for recall, the sums were found to be 198 cases for anticipation, and 146 for study–test. This seems to indicate that, (*a*) associative learning between the first instance of the concept and its position or stimulus number was greater or better during anticipation learning than during study–test learning, or (*b*), showing the *S*

the response terms during recall allowed him to "deduce" the subsequent occurrence of the first instance of a different concept. Both factors may be involved, although the latter seems more reasonable. Finally, the data in Table 1 are clear in demonstrating that failure of *S* to recall the first instance of a concept provides very little penalty under the anticipation procedure, since he can proceed to give other instances of the concept for the two following positions.

The evidence seems to point to the fact that learning under the study–test method is relatively fragile or weak with regard to positions or stimulus numbers which mark the first instance of a new concept. The same is probably true for the anticipation learning. However, on the recall trial the *S* having learned by the study–test method receives no feedback information and he has few means to apply corrective procedures based on any knowledge he had about the conceptual structure. In effect, he was reduced to responding on an item by item basis, much as is the case for an unstructured list. On the other hand, the *S* having anticipation learning could apply corrective procedures based on his knowledge about the conceptual structure of the list.

General Discussion

The higher the conceptual organization of the lists the more rapid was the learning. It is presumed that this relationship was produced because the greater the number of the conceptual levels the more precise was the placement of an item if the rules indicated by the conceptual relationships were followed. Nevertheless, this should not be taken to mean that the hierarchical structure in its totality entered into the learning of the most completely structured lists. The *S* could learn the sequence of eight concepts of three instances each without reference to the higher-order concepts present in the list (animals, plants, beverages, minerals). So also, List 3 could be learned by reference only to the four successive concepts of six instances each and without the living–nonliving distinction per se entering into the learning. The error data were clear in showing that at least one level of conceptual responding was involved in the learning of Lists 3, 4, and 5, but these error data do not speak to the question of whether two or more conceptual levels were involved in learning Lists 4 and 5. To determine if two or more levels were involved would require the use of a list in which the blocks of three instances were ordered randomly with respect to the more inclusive concepts. A list of this type was not included in the present study.

In the previous study (Underwood & Zimmerman, 1973), learning was facilitated up through the structure corresponding to List 3 in the present study. No further enhancement occurred for Lists 4 and 5. This is believed due to the fact that the present lists were made up of more obvious concepts and more obvious concept instances than were the previous lists. Unlike the present findings, frequency of the concept instance with a concept did not influence learning in the previous experiment. In view of the findings of the auxiliary experiment, it seems now that this contradiction is another by-product of the differences in the learning by the two methods.

The major purpose of the present study was to attempt to identify the component(s) in memory which had led to the increase in recall over 24 hr. as list structure increased. In an inelegant way, this study was quite successful in achieving its purpose. The use of the study–test method of learning was followed by a complete lack of any differences in the recall tests, whether paced or unpaced. Only relearning, reflecting the same relationship as was found in original learning, was related to list structure. With the study–test method the original learning was apparently based on very weak associations between stimulus number or position and words marking the conceptual changes which occurred throughout the list. The loss of these associations over 24 hr. essentially made the conceptual nature of the lists useless as a recall vehicle. The items which were recalled were apparently based upon factors specific to them.

Considering now the broader context of these studies, the question may be raised about the systematic importance of the earlier study (Underwood & Zimmerman, 1973), and the auxiliary experiment of the present report, in both of which the anticipation method was used. Recall and list structure were directly related in these studies. However, the error data from this previous study gave no support to the idea that forgetting of unstructured lists was produced by interference among the con-

ceptual associations among the words. In the present study there was no evidence that conceptual associations played any interfering role in learning the unstructured lists. It does not seem now, therefore, that these types of lists have any special or pointed use for the study of interference as a source of forgetting.

REFERENCES

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psych[illegible] Monograph*, 1969, **80**(3, Pt. 2).

BOUSFIELD, W. A., & COHEN, B. H. M[illegible]linity-femininity in the free recall of a catego[illegible] stimulus word list. *Perceptual and Motor* [illegible] 1956, **6**, 159-166.

UNDERWOOD, B. J. Degree of learni[illegible]d the measurement of retention. *Journal* [illegible] *Verbal Learning and Verbal Behavior*, 1964, 3 [illegible]129.

UNDERWOOD, B. J., & ZIMMERMAN, J. [illegible] retention as a function of hierarchical struc[illegible] *Journal of Experimental Psychology*, 1973, 9[illegible] 242.

(Received August 31, 1973)

Jour ...ental Psychology
19.. , No. 5, 863 867

CHANGE IN SPEAKER'S VOICE AND RELEASE FROM PROACTIVE INHIBITION

JOHN M. GARDINER[1] AND PAULINE C. CAMERON[2]

The City University, London, England

Four triads of unrelated words were presented auditorily in the Brown–Peterson distractor task. Over Trials 1 to 3, half of the *S*s heard the words spoken in a female voice and half in a male voice. On Trial 4, *S*s in a control group heard the words in the same voice as on preceding trials and *S*s in an experimental group heard the words in the alternate voice. Proactive inhibition (PI) developed rapidly over Trials 1 to 3 in both groups. On Trial 4, however, *S*s in the experimental group recalled significantly more words than did *S*s in the control. It was concluded that *S*s can utilize change in speaker's voice to reduce the cumulative effects of PI in short-term memory.

Considerable interest has been shown recently in the investigation of those attributes or dimensions of verbal items which *S*s can utilize or identify in a variety of memory tasks (Underwood, 1969; Wickens, 1970). One strategy, which has yielded a considerable body of evidence on the encoding of various attributes of verbal items, has been the use of the release from proactive inhibition (PI) technique. This technique capitalizes upon the rapid buildup of PI which occurs over the first few trials in the Brown–Peterson distractor task (Keppel & Underwood, 1962) to determine the extent to which *S*s can utilize a specific change in the nature of the to-be-remembered material to reduce the cumulative effects of PI. Wickens (1970, 1972) has reviewed many studies which demonstrate that when the "category" of the material is changed after three or four trials, recall will be better than if the category remains unchanged. Wickens has argued that the amount of recovery in performance (or release from PI) on the critical trial, as a function of any particular "category shift," provides an index of the relative effectiveness of that category as an attribute or encoding dimension of memory.

A second strategy has been to examine more directly *S*s' retention of information about various attributes of verbal items, as, for example, in determining the extent to which *S*s can identify input modality following recall of a list of words presented in mixed modes (Bray & Batchelder, 1972). In this situation, there is evidence not only that *S*s are able to make accurate judgments concerning between-modality attributes (whether an item was presented visually or auditorily), but also concerning within-modality attributes (whether an item was presented in upper- or lowercase letters, or in a male or female voice). Hintzman, Block, and Inskeep (1972), for example, gave *S*s an unexpected final recognition test following several free-recall trials in which the word lists were presented in one of three mixed modes: visual or auditory; upper- or lowercase; male or female voice. In all cases, the results showed that, given recognition of the item in the final test, *S*s could identify the input mode with considerable accuracy for at least several minutes after presentation. In a rather different vein, Craik and Kirsner (in press) investigated the persistence of information about speaker's voice (male or female) using the continuous recognition paradigm of Shepard and Teghtsoonian (1961). Craik and Kirsner found that when test words were represented in the same voice as in the initial presentation, recognition performance was better than when the second presentation was in a

[1] Requests for reprints should be sent to John M. Gardiner, Department of Social Science and Humanities, The City University, St. John Street, London EC1V 4PB, England.

[2] Now at the Department of Psychology, Brunel University, Uxbridge, Middlesex, England.

different voice. This "same voice" facilitation did not decline over a 2-min. lag.

There is thus good evidence that *Ss* retain various aspects of input mode information for at least several minutes following item presentation, and also that such information may in certain circumstances facilitate memory performance. These results are important in that they imply that the representational characteristics of verbal material may persist for much longer than many theorists have maintained (e.g., Crowder & Morton, 1969).

Such evidence suggests also that *Ss* might be able to utilize input modality attributes to reduce the cumulative effects of PI in the Brown–Peterson distractor task. Indeed, several recent studies (Hopkins, Edwards, & Cook, 1973; Kroll, Bee, & Gurski, 1973) have established that between-modality shifts (visual to auditory or auditory to visual presentation) can lead to dramatic release from PI effects. The results of several within-modality shift studies have also been reported. Within the visual modality, Shearer and Wickens (unpublished study cited by Wickens, 1972) found that a shift from lower- to upper-, or upper- to lowercase letters did not produce a significant release effect. Shifting slide background area (Turvey & Egan, 1969), and the color of the slide background (Reutener, 1972), were, however, found to be effective. Somewhat surprisingly, the effect of shifts within the auditory modality have not hitherto been reported. Bearing in mind the results obtained by Craik and Kirsner (in press) and Hintzman et al. (1972), it seems quite possible, for example, that change in speaker's voice would lead to a release from PI effect. The present study was designed to investigate this possibility.

Method

Design. All *Ss* received four triads of unrelated words presented auditorily in the Brown–Peterson distractor task. There was an experimental and a control group of *Ss*. Over Trials 1 to 3, half of the *Ss* in each group heard the words spoken in a female voice and half in a male voice. On Trial 4, *Ss* in the control group heard the words spoken in the same voice as on previous trials. The *Ss* in the experimental group heard the words spoken in the alternate voice on Trial 4.

Subjects. The *Ss* were 120 volunteers, mostly male undergraduate students, who were paid for their services. Sixty *Ss* were assigned randomly to each group and the *Ss* were tested individually.

Materials. Thirty unique 12-word sets were constructed by sampling without replacement from a pool of 600 common two-syllable nouns. Each set was divided arbitrarily into four word-triads. The word sets were yoked across experimental condition such that each set was used four times once for each voice condition within each group

Procedure. The word lists were recorded on a Brenell STB 2 tape recorder and subsequently presented to *S* via a loudspeaker. Each *S* was tested in a small soundproof cubicle. The trial procedure was as follows. Each trial began when the tape recorder was switched on and *S* heard a set of three words presented at the rate of one word per sec. The tape was then switched off and the "click" made by the switch was the *S*'s cue to begin the distractor task. This task was that of reporting aloud the sum of each of a series of pairs of random single digits from a sheet of these placed on a table in front of the *S*. The addition task was paced at a 1-sec. rate by the click of a metronome. After 18-sec. addition, *E* signaled with a pencil tap that *S* was to attempt oral recall of the three to-be-remembered words. The *S* was allowed 9 sec. for recall, after which a further pencil tap signaled that the next set of words was about to be presented. There was no intertrial interval, and each complete trial sequence lasted for 30 sec.

All *Ss* were instructed that the experiment involved two "unrelated" tasks, an arithmetic task and a memory task, and that performance on both tasks would be scored. The *Ss* were given an initial practice at the distractor task alone. They were then given seven lead-in trials presented in the same manner as described above. Lead-in trials comprised triads of one-syllable common words and all *Ss* received the same words. Prior to the lead-in trials the *Ss* were informed that the words would sometimes be spoken in a male voice and sometimes in a female voice, that changes in voice occurred at "random," but that all they had to do was to remember the words. The seven lead-in trials were in fact a random sequence with respect to voice. Between the lead-in trials and the four experimental trials there was an unfilled interval of about 2 min.

Results

Preliminary analyses indicated that there were no reliable differences as a function of voice (male or female) and the data were pooled over this variable. Figure 1 shows the percentage correct recall for the experimental and control groups. One point was given for each word recalled correctly, re-

gar...ss of its output position. The figure sh... that recall declined rapidly in each ...ver the first three trials. On the ... trial, however, the experimental ... showed a substantial recovery in ...mance, whereas performance in the ...ol group remained essentially the same as ... was on Trial 3.

...parate analyses were carried out on d... a from the first three trials and on data f...m Trial 4. Considering Trials 1 to 3, t... results of an analysis of variance showed ... the only significant effect was that due ... trials, F (2, 236) = 33.15, $p < .001$. ...he F ratios associated with groups, and ...oups × Trials, were less than unity. ...hese results indicate that the build-up of I over Trials 1 to 3 was reliable and did not differ with respect to groups. The result of an uncorrelated t test carried out on data from Trial 4, however, confirmed that the superiority shown by the experimental group was highly significant, t (118) = 3.95, $p < .001$.

Following the procedure suggested by Wickens (1970) the amount of release from ...I was assessed by dividing the difference between the experimental and the control ...oups on Trial 4 by the amount of loss shown by the control group from Trials 1 to 4. Use of this procedure gave a percentage recovery of 67%. Relative to previous findings (Wickens, 1972, Figure 3), this represents a substantial amount of release.

After recall of the final triad, each S was asked which voice or voices had been used over the last four trials. Somewhat surprisingly, only 33 Ss in the control group, and 29 Ss in the experimental group, reported accurately the voice in which the *final* triad had been presented. Since there were 60 Ss in each group, these figures are very close to a chance level of performance, and many Ss commented that they paid no attention to voice, or were guessing when they reported on which voices they had heard. In this context it is of interest to note how Ss' reported awareness of the change in voice relates to the release effect. Of the 60 Ss in the experimental group 22 reported the change in voice from Trials 3 to 4 accurately. Taking an increase in the number of words recalled by the S from Trials 3 to 4 as an index of "individual release," 12 of the 22 Ss who noticed the change and 16 of the 38 Ss who did not notice the change showed individual release effects. These results imply that the recovery in performance shown by Ss in the experimental group on trial four did not depend upon Ss' awareness of the change in voice.

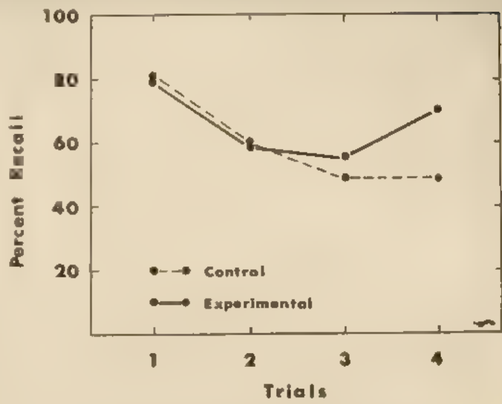


FIGURE 1. Percentage correct responses for control and experimental groups.

DISCUSSION

Clearly, the results of the present study show that change in speaker's voice lead to a dramatic recovery from the effects of PI in the Brown–Peterson distractor task. Thus speaker's voice may be added to the already large list of attributes which Ss apparently encode in short-term memory for verbal items.

The magnitude of the release effect obtained runs counter to the general finding that "physical," as opposed to semantic shifts, do not typically produce large release effects (Wickens, 1972). Why should change in speaker's voice be an exception to the more general result? One possibility is that the Ss in the present study were "set" to encode items in terms of voice. That is, the Ss were informed beforehand that the triads might be spoken in either a male or a female voice and received several lead-in trials in which they heard each of the two voices. In most previous release from PI studies, with the exception of some between-modality shift experiments (e.g., Hopkins et al., 1973), the Ss were not informed beforehand that any change in the

nature of the to-be-remembered material would occur.

However, as Underwood (1972) has pointed out release effects in general may to some extent reflect "priming" the *S*, implicitly if not explicitly, to encode the material along the specific dimension being investigated. We may well be "overloading memory" if we assume that *S*s habitually or automatically encode verbal materials along all the many dimensions so far shown to yield release effects. The more cautious conclusion, of course, is simply that *S*s can utilize such dimensions to reduce the effects of PI when given the "opportunity" to do so, and, therefore, that the dimensions may be represented, in some sense, by different memory codes. In view of Underwood's point, it may be worth emphasizing that this conclusion is independent of any conclusion with respect to the possible influence of priming the use of particular memory codes. Underwood's argument implies, rather, that the employment of such codes may, to a large extent, be "optional" rather than "automatic" (cf. Wickens, 1970). If this be so, then an urgent task for further research will be that of investigating the conditions that tend to favor the use of one encoding strategy rather than another. Reutener (1972), for example, found that shifting the color of the slide background was effective in producing release from PI when the stimulus materials were arabic numbers or spelled out numbers, but not when the materials were words drawn from a taxonomic category. On the basis of this result, Reutener suggested that a "physical" shift may be effective only when the to-be-remembered material is not words. While such a conclusion is also in line with results reported by Turvey and Egan (1969) and Shearer and Wickens (unpublished study cited by Wickens, 1972), the results of the present study do not fit into this pattern. The results of the present study, rather, suggest that *S*s may also place more reliance on "physical" codes when the to-be-remembered material consists of relatively "unrelated" words. Nevertheless, one plausible factor which may affect the extent to which *S*s utilize "physical" forms of coding may well be the degree of meaningfulness which *S*s can extract from the stimulus materials.

That the *S*s in the present study were not able to report accurately which voices they had heard in the final trials is somewhat surprising in view of the results obtained by Craik and Kirsner (in press). These authors found that *S*s could assign speaker's voice with 65–75% accuracy after a 2-min. interval. This discrepancy may be related to the fact that whereas *S*s in the Craik and Kirsner study identified presentation voice in the context of the presented words, *S*s in the present study were required to recall presentation voice in the absence of the words. We suggest that the inability of our *S*s to report accurately which voices they had heard in the final trials is in good agreement with the notion that the auditory information in speaker's voice was retained in some literal or representational form. Our data may thus be viewed as adding to the already substantial body of evidence (Craik & Kirsner, in press; Hintzman et al., 1972) which indicates that auditory information may persist for much longer than many theorists have maintained (e.g., Crowder & Morton, 1969). Further, the present study extends these findings by demonstrating that such information can to a large extent eliminate the effects of PI in short-term memory.

REFERENCES

Bray, N. W., & Batchelder, W. H. Effects of instructions and retention interval on memory of presentation mode. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 367–374.

Craik, F. I. M., & Kirsner, K. The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, in press.

Crowder, R. G., & Morton, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, **5**, 365–373.

Hintzman, D. L., Block, R. A., & Inskeep, N. R. Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 741–749.

Hopkins, R. H., Edwards, R. E., & Cook, C. L. Presentation modality, distractor modality, and proactive interference in short-term memory. *Journal of Experimental Psychology*, 1973, **98**, 362–367.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Kroll, N. E. A., Bee, J., & Gurski, G. Release of proactive interference as a result of changing presentation modality. *Journal of Experimental Psychology*, 1973, **98**, 131–137.

Reutener, D. B. Background, symbolic, and class shift in short-term verbal memory. *Journal of Experimental Psychology*, 1972, **93**, 90–94.

Shepard, R. N., & Teghtsoonian, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, **62**, 302–309.

Turvey, M. T., & Egan, J. Contextual changes and release from proactive interference in short-

...rbal memory. *Journal of Experimental ...ogy*, 1969, **81**, 396–397.

...ood, B. J. Attributes of memory. *Psy...al Review*, 1969, **76**, 559–573.

...wood, B. J. Are we overloading memory? ...A. W. Melton & E. Martin (Eds.), *Coding ...esses in human memory.* Washington: V. H. ...inston, 1972.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

Wickens, D. D. Characteristics of word encoding. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington: V. H. Winston, 1972.

(Received September 14, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 868–874

THE COMBINED EFFECTS OF STIMULUS AND RESPONSE CONDITIONS ON THE DELAY IN IDENTIFYING THE PRINT COLOR OF WORDS[1]

RAYMOND L. MAJERES[2]

Western Illinois University

The interaction between stimulus and response conditions in the Stroop interference effect was studied in two experiments. Response conditions in Experiment I required college students ($N = 36$) to identify colors by using verbal responses (color names, color associates, or single letters). In Experiment II, *Ss* ($N = 36$) identified the colors by writing the first letter of the responses used in Experiment I. Stimulus conditions in both experiments were color only, color names in nonmatching print colors, and color-associated words in unrelated print colors. In line with predictions from Morton's word recognition model, the Stimulus × Response interaction was significant in both experiments ($p < .01$). Differences between the experiments indicate that the type of response (verbal or motoric) influences the delaying effect of the words.

Naming the color of printed words (e.g., such as *friend*, *sky*, or *run* printed in red) takes more time than naming the color of colored patches (Fox, Shor, & Steinman, 1971; Klein, 1964). This effect is especially strong when the words are color names (Stroop, 1935). It is as if the highly developed systems involved in reading operate autonomously and interfere with simple color naming.

Klein (1964) used different kinds of verbal material in addition to color names as potential sources of distraction. It was found that a variety of verbal materials, including nonsense syllables, could delay color naming. The nature of the words in terms of their semantic relationship to the colors (other color names, color-related words) and their familiarity (word-count frequency) influenced the degree of delay. These particular results have been replicated, and a variety of word characteristics have been found to influence the delay (Dyer, 1973). Both Stroop (1935) and Klein emphasized the role of response competition, involving central as well as peripheral aspects. Implicit in their emphasis was the initial parallel processing of both color and word, with interference arising from difficulties in response selection (see also Dyer, 1973).

Studies by Treisman and Fearnley (1969) and Hock and Egeth (1970) have questioned this assumption, pointing out that the task as used confounded stimulus and response effects. They attempted to eliminate this confounding by devising tasks that did not require the overt naming of colors, but rather involved judgments of perceptual match or a match with "target" stimuli held in memory. Treisman and Fearnley suggested that the critical feature of the Stroop effect is that a comparison must be made across hypothesized word and color "analyzers," and such comparisons are less efficient than those comparisons made within an analyzer. Also, "across-analyzer" comparisons were believed to be more susceptible to interference effects from values of the irrelevant attribute. However, Egeth, Blecker, and Kamlet (1969) found that an interference effect can be obtained in "within-analyzer" comparisons. In their task the responses were verbalized and the response words (*same* and *different*) were printed in the different ink colors. That is, the identity of the set of words used in the stimuli to the set of response terms did produce a delay, even though the task involved within-attribute matching.

A few other studies have provided data

[1] This research was carried out while the author was a postdoctoral fellow supported jointly by Educational Testing Service and National Institutes of Health Grant MH-08260.

[2] Requests for reprints should be sent to Raymond L. Majeres, Department of Psychology, Western Illinois University, Macomb, Illinois 61455.

on the mutual effects of both stimulus and response variables. Most of these studies were designed to investigate language organization in bilinguals (Dalrymple-Alford, 196… Preston & Lambert, 1969). Preston and Lambert considered initial identification as one of the factors in the observed interference effect, but output processes were considered to be the major factor in producing the interference effect. Dalrymple-Alford, in a study using Arabic–English bilinguals, introduced the possibility that the stimulus word primes all words associatively related to it. That is, not only does the stimulus word prime the wrong response, but it also primes associatively related responses. There is more recent evidence that this priming may extend beyond simple associational processes and probably involves the word's semantic category as well (Warren, 1972). This is in contrast to Hock and Egeth's (1970) interpretation, where the responses were considered to sensitize *S*s to the stimulus word. It may be that the stimulus word and the response both alter the availability of some central linguistic unit.

Morton (1969) has suggested that the availability of such mediational units, called logogens, is affected by stimulus information (seeing a color, seeing a word, hearing a word, etc.) and by the experimental context. In the Stroop (1935) paradigm, the experimental context refers primarily to the instructions to identify colors, which presumably increases the availability of (primes) color names, color-related words, and perhaps entire color categories (Warren, 1972). A third factor affecting the availability of the logogen to a serial output buffer is the response. Response repetition (overt or covert) is believed to lower the thresholds of the respective logogens. Hence, in the Stroop task, both stimulus and response conditions should influence logogen availability.

For example, the use of color names both as stimulus words (printed in incongruent colors) and as response terms (as names of the print colors) should provide the greatest competition in terms of the greatest number of incompatible logogens reaching threshold: color identification is required, which influences availability of color names in general; color names are irrelevant sources of stimulus information, which increases the availability of incompatible color names; and color-name responses are required, which lowers the thresholds for color names. Likewise, using color associates as response terms, such as saying *fire* for red, should have a direct priming effect on color associates and not on the color names. If color associates are in the stimulus card as irrelevant, incompatible information and are also used as responses, the logogens for color associates should be highly available (via stimulus information), and their thresholds should be lowered (via response requirements). Such conditions should produce a dramatic delaying effect analogous to the original Stroop (1935) task. However, since the context is still color identification, which has its primary effect on the availability of color names rather than of color associates, the effect may be somewhat reduced.

Using single letters as vocal responses should not alter the threshold for color-associate or color-name logogens. However, stimulus words (color names and color associates) and context conditions should produce delaying effects similar to those found in the original Stroop (1935) task, with color names producing more delay than color associates. But the magnitude of the delay may be reduced, since single-letter responses should not alter logogen thresholds, except possibly by verbal mediation.

Experiment I

An initial experiment was designed to demonstrate the joint influence of both stimulus and response conditions on the delay in color naming, using procedures directly comparable to those used by Stroop (1935). Both the kind of words in the stimulus (color names, color-related words, and color only with no verbal distraction) and the kind of color-identifying responses (conventional color names, color-associated words, or arbitrary letters) were

TABLE 1

COLOR IDENTIFICATION TIMES (IN SEC.) FOR THREE RESPONSE GROUPS ON THREE STIMULUS CONDITIONS: EXPERIMENT 1

Type of interference	Trials				*M*	[illegible]rease [illegible]ntrol
	1 & 2	3 & 4	5 & 6	7 & 8		
Response Group N						
None	19.6	19.9	19.2	18.8	19.1	
Color associates	27.3	26.1	25.4	25.2	26.0	[illegible].5
Color names	37.4	35.6	32.8	32.6	31.6	[illegible]
Response Group A						
None	34.2	33.5	33.7	31.4	33.2	
Color associates	47.7	46.9	43.3	41.9	45.0	[illegible]5
Color names	46.2	45.0	43.0	40.2	43.6	[illegible]4
Response Group SL						
None	41.1	38.3	36.0	34.6	37.5	
Color associates	48.5	43.4	39.7	40.0	42.9	[illegible]
Color names	54.4	49.7	43.4	41.0	47.1	[illegible]

Note. Abbreviations: N = color names, A = color-color associates, SL = single letters.

systematically varied to investigate their joint influence.

Method

Subjects. Thirty-six students (18 males and 18 females) from a private and a state college in New Jersey volunteered to take part and were paid $2 for a 1-hr. session.

Materials. Three sets of stimulus cards were prepared, as described below.

Card 1. The words *red, green, blue,* and *orange* were printed in lowercase letters (pica) by a multilith-offset process on 8½" × 11" sheets with eight rows of five words each. The words were printed in incongruent ink colors (words never congruent with ink color) with each color appearing 10 times in a quasirandom order (doublets were avoided within rows). Each word appeared about the same number of times (4 times) in each of the three possible incongruent colors. This was the color-names stimulus condition.

Card 2. Using the same format as in Card 1, the words *fire, sky, grass,* and *juice* were printed in colors incongruent with the colors suggested by these words: red, blue, green, and orange, respectively. This card was the color-associates stimulus condition. The color-associated words and the color-names were quite comparable in terms of Thorndike–Lorge (1944) word-count frequency, with all rated AA except *orange* (A) and *juice* (37).

Card 3. Sets of five asterisks (approximating the average length of the words) were printed with each set in one of the four colors in the same format as was used on Cards 1 and 2. This was a noninterference or color-only stimulus condition, in that no words were printed on this card. There were two different forms of each of these three stimulus cards.

Procedure. Equal numbers of male and female *S*s were assigned to three different experimental groups. In the first group, Group N, *S*s [illegible] the colors on the three stimulus cards usi[illegible]ventional color names (red, blue, green, and orange). In the second group, Group A, *S*s identified the colors by using color-associated words, saying *fire* for red, *sky* for blue, *grass* for green, and *juice* for orange. In the third group, Group SL, the identifying responses were single letters: *v* for red, *x* for green, *y* for blue, and *z* for orange. All *S*s were given practice trials in serially naming the 40 color-only items using their particular set of identifying responses to a performance criterion of 60 sec. This practice was designed to ensure that the responses were sufficiently familiar to permit smooth and accurate responding. No *S* in the N or A groups required more than two practice trials to reach criterion. Of course, *S*s in the single-letter (SL) group typically required more extensive practice to reach the criterion since the labels they used were novel and arbitrary.

After the initial training trials, *S*s in all three response groups identified colors on the three different stimulus cards. There were eight trials on each of the three stimulus conditions. Presentation of each of the three stimulus cards constituted a trial. The color-only stimulus condition always preceded the other two stimulus conditions on each trial. The order of the color-associates and color-names cards was alternated between trials. Half of the *S*s began with a color associates-color names order, and half began with the reverse order.

A stopwatch was used to obtain performance times, and a 30-sec. interval between cards provided *S*s with a brief rest and enabled the recording of *S*s' performance times. The *S*s were instructed to correct all errors so that any errors would be reflected in the time score. The number of uncorrected errors was too limited to permit a meaningful analysis.

Res

Pe ·rmance times on each stimulus card were ·nalyzed using a 3 × 3 × 2 × 8 analysis ·· variance design with response conditio· (color names, color associates, and sin· · letters), stimulus conditions (color na· ·es, color associates, and color only), se· and trials (eight) as factors. Response c·· ·litions and sex were between-*S*s variab· ·s, and stimulus conditions and trials · within-*S*s variables.

·lean times to identify the colors under · different experimental conditions are ·ven in Table 1. Response conditions, F ·, 30) = 17.01, $p < .001$; stimulus condi·ons, F (2, 60) = 190.55, $p < .001$; and ·heir interaction, F (4, 60) = 13.48, $p <$ .001, were significant. There were significant trial effects, F (7, 210) = 28.11, $p <$ .001; as well as significant Trial × Response, F (14, 210) = 4.48, $p < .001$, and Trial × Stimulus, F (14, 420) = 4.56, $p <$ .001, interactions. Neither main effect for sex nor any of the other interactions were ·ignificant. Since the Stimulus × Re·ponse interaction was significant, the sig·ificant main effects are interpreted within ·at context.

Stimulus × Response interaction. The prediction of a Stimulus × Response interaction was confirmed. The degree of delay in color naming did depend on the particular combination of stimulus and response terms (Figure 1). Tests of simple effects indicated that the two interference conditions (color names and color associates) did produce significant delays in color identification, and this was the case for each of the three response groups (Newman–Keuls, ps $< .01$). Differences between the color-names and color-associates stimulus conditions were significant (Newman–Keuls, ps $< .01$) for both the color-names response group (34.6 vs. 26.0) and for the single-letters response group (47.1 vs. 42.9), but this difference was not significant for the color-associates response group (43.6 vs. 45.0).

Practice effects. With respect to the different response conditions, the practice effects were more marked for the newly learned conditions (identifying colors with color associates or with arbitrary single letters) than for the already well-learned color-names response condition (Table 1).

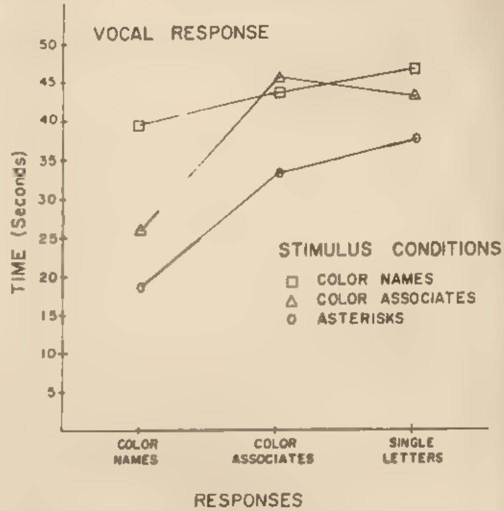


FIGURE 1. Mean vocal identification times for three stimulus conditions by three different response groups.

The effects of practice were much greater for the two interference conditions than for the color-only condition. An analysis of variance with response conditions and sex collapsed indicated that this effect was reliable, for the Stimulus × Trials interaction, F (7, 245) = 7.92, $p < .001$. Practice not only improved the speed of color identification per se, but also improved the efficiency with which the delay produced by the words was overcome. However, the nature of the Stimulus × Response interaction was not significantly affected by practice.

EXPERIMENT II

Pritchatt (1968) argued that the Stroop (1935) effect is primarily a verbal phenomenon. Using a simple motor response, tapping colored keys, it was found that the magnitude of the Stroop effect could be reduced but not eliminated. Pritchatt reasoned that sufficient practice may eliminate the Stroop effect by reducing the role of verbal mediation.

Also, according to Morton's (1969) model, vocal responses both increase the word's availability because of auditory stimulus

TABLE 2

Color Identification Times (in sec) for Three Response Groups Under Thr[illegible] Stimulus Conditions: Experiment II

Type of verbal interference	Trials				M	[illegible]crease [illegible]ontrol
	1 & 2	3 & 4	5 & 6	7 & 8		
Response Group N						
None	28.4	27.2	25.8	25.0	26.6	
Color associates	36.7	32.7	31.0	29.6	32.5	[illegible]2.1
Color names	41.9	39.0	34.6	33.8	37.3	[illegible]0 2
Response Group A						
None	37.1	36.9	35.5	32.3	35.5	
Color associates	54.9	44.9	43.8	41 3	46.2	[illegible]0.2
Color names	53.5	46.5	42.6	40.9	45.9	[illegible]9.4
Response Group SL						
None	41.2	38.8	35.4	35.5	37.7	
Color associates	46.6	42.7	40.3	36.5	41.5	[illegible]0.1
Color names	57.4	48.2	44.5	39.5	47.4	[illegible].7

Note. Abbreviations: N = color names, A = color associates, SL = single letters.

information and lower the word's threshold because of response repetition. A written and abbreviated response, such as the first letter of each of the responses of the first experiment, should eliminate the auditory stimulation and also eliminate other feedback from overt speech mechanisms, replacing them with a modest amount of visual and kinesthetic information.

Furthermore, as suggested by Pritchatt (1968), using a simple, mediated (by color names) motor response may result [illegible] a reduction or elimination of the medi[illegible]g role of the color name with practice. T[illegible]esult of such conditions should be a red[illegible]on in the interference effect. However, [illegible] context, the requirement for color id[illegible]fication, would be unchanged, and covert responding also should lower word thresholds (Morton, 1969). The net result should be a general pattern of results similar to those obtained in Experiment I (a significant Stimulus × Response interaction), however, practice may eliminate the Stroop effect by the elimination of verbal mediation. In this experiment, simple verbally mediated "motor" responses replaced the vocal responses of Experiment I. The design was analogous to the first experiment except for the response requirement.

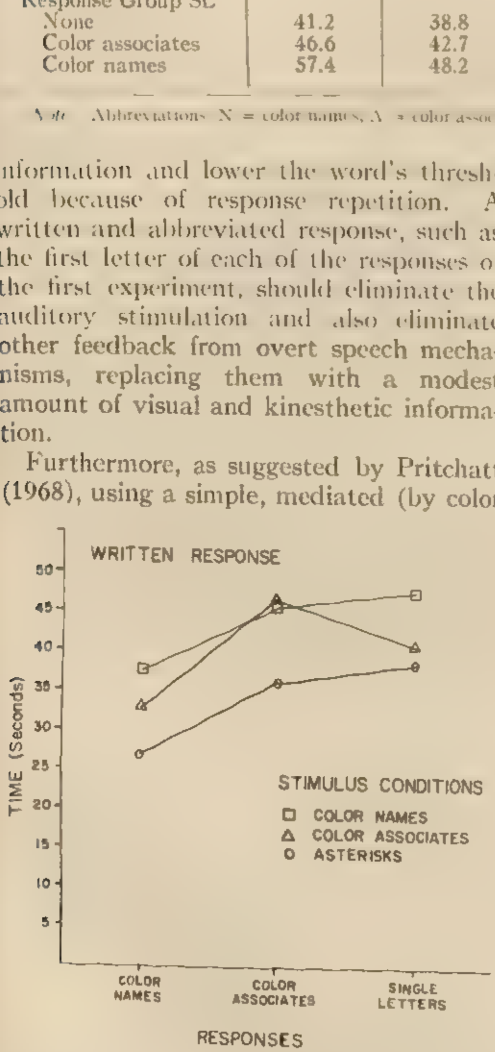


Figure 2. Mean written identification times for three stimulus conditions by three different response groups.

Method

Subjects. Thirty-six volunteers (18 males and 18 females) from a New Jersey state college served as *Ss* and were paid $2 for a 1-hr. session.

Procedure. The major modification of the procedure was to require a written response rather than a vocal one. This was done by having *Ss* in three response groups write the first letter of the identifying response that was used in the first experiment below each item.

All *Ss* were given training trials where they vocalized the correct responses just as in the first experiment. The *Ss* then were given one written practice trial of 40 responses to colors only, and then the entire experimental series was given. Each *S*

was gi... ne sheet at a time and was timed from the fir... o the last response. Errors were to be correc... by crossing out the incorrect letter. Uncorrec... errors, as in the first experiment, were rare a... insufficient for a meaningful analysis. All *Ss* w... tested individually.

R...

... in the first experiment, a 3 × 3 × 2 ... analysis of variance design was used. ...n times for the different experimental c...ditions are presented in Table 2. The ...lysis of performance times yielded sig...icant main effects for the stimulus con...tions, F (2, 60) = 101.05, $p < .001$; re...ponse groups, F (2, 30) = 7.02, $p < .01$; ...nd trials, F (7, 210) = 53.78, $p < .001$. A significant Stimulus × Response interaction was found, F (4, 60) = 5.00, $p < .01$; as well as significant Stimulus × Trials, F (14, 420) = 7.51, $p < .001$, and Response × Trials, F (14, 210) = 1.89, $p < .05$, interactions.

Unlike the first experiment, there were significant Stimulus × Response × Trials, F (28, 420) = 1.94, $p < .01$, and Stimulus × Response × Sex, F (4, 60) = 3.73, $p <$ ...01, interactions. This latter interaction was due to females in the color-names response group doing significantly better than females in the single-letters response group on the color-names stimulus condition, F (1, 30) = 11.45, $p < .01$; this was the only such difference to reach statistical significance. The sex main effect and all other interactions were not significant.

Stimulus × Response × Trials interaction. The significant Stimulus × Response × Trials interaction was the result of a general decrease in the magnitude of the differences among the groups over trials (Table 2). The effects of practice did not eliminate the delay produced by the words by the final trial (Trial 8) for the color-names (first letter of color name written) group (Newman–Keuls, $p < .01$). However, the delay produced by the color associates was not significant on the last trial. Similar results were found for the single-letters group. There was still a significant delay in color identification for the color-names stimulus condition (Newman–Keuls, $p < .05$), but the delay was nonsignificant for the color-associates stimulus condition. For the color-associates response group, there were significant interference effects for both stimulus conditions (Newman–Keuls, $p < .01$).

Stimulus × Response interaction. Again, as in the first experiment, the predicted interaction of stimulus and response conditions was found (Figure 2). Tests of simple effects indicated that the two interference conditions (color names and color associates) again produced significant delays in color identification for each of the three response groups (Newman–Keuls, ps $< .01$).

Differences between the color-names and color-associates stimulus conditions were significant (Newman–Keuls, ps $< .01$) for both the color-names and for the single-letters response groups, but this difference was not significant for the color-associates response group.

Discussion

Interference effects. For each of the three vocal response groups of Experiment I, times to identify colors on the color-word cards were significantly greater than on their respective color-only cards. Hence, the words significantly increased the time required to identify the colors under all three response conditions. An interference effect was obtained independent of the type of identifying response. However, the amount of interference appeared to be related to the type of response and to the type of verbal material on the stimulus card.

The results from the second experiment were remarkably consistent with those of the first experiment. Again, the stimulus words (color names and color associates) produced significant delays in color identification for all three response groups. The delaying effects of color associates, first reported by Klein (1964), and color names were replicated and extended to several other response conditions.

Maximum delay in color naming (78% increase over control conditions) was found when the distracting stimulus words were color names and the response terms were also color names. Similarly, in Experiment II, the analogous conditions (color-names stimulus conditions and color-names mediated response conditions) produced the greatest amount of delay, but the magnitude of the effect was reduced (only 40% increase over control condition).

The next most interfering condition was expected to be the color-associates response group under the color-associates stimulus conditions (Experiment I). This set of conditions was found to produce about the same delay as the color-names stimulus condition when using color associates as responses (color associates, 35.5% vs. color names, 31.4%). In Experiment II, the color-associates stimulus condition using the color-associates mediated responses produced similar results (30.2% vs. 29.4%). Although the effect of having color associates in both stimulus and response conditions did not result in the pronounced effect characteristic of the Stroop conditions, the effect was large enough to eliminate the usually observed difference between color-names and color-associates stimulus conditions. The results indicate that the traditional Stroop (1935) phenomenon is a product of both the particular stimulus words and the response terms.

Practice. Practice in both experiments reduced the delay in color identification produced by the words. There was improvement in color identification per se, but, in addition, there was improvement in overcoming the interference. In the color-names response condition of Experiment II, only the color-names stimulus condition still produced significant interference by the final trial. The delay for the color-associates stimulus condition was not significant. A similar result was found for the single-letters response group. This result is interpreted as evidence that context, vocal color identification, maintained the availability of color names in some sustained fashion, which is a feature of Morton's (1969) "context system."

The spoken responses in Experiment I, in contrast to Experiment II, did not result in comparable practice effects. Presumably, the activation or involvement of the speech system resulted in an attentional bias for verbal stimuli which was difficult to overcome and which produced significant interference for color-names and color-associates stimulus conditions for all three response groups. While in Experiment II, with the motor response of writing, only the delay produced by the color names resulted in sustained delaying effects, probably due to the color identification context. Since the color-associates stimulus condition only produced a sustained delay in the color-associates mediated response condition, this suggests that the thresholds of color associates were lowered by this response requirement, thereby preventing the elimination of color-associate mediation with practice. This interpretation is supported by the finding that color-associates produced no significant delay (Trial 8) in the other two response conditions, color-names and single-letters conditions.

The results support the interpretation of the Stroop (1935) phenomena in terms of verbal mediation processes (Pritchatt, 1968) which are affected by experimental context, stimulus priming (Dalrymple-Alford, 1968; Morton, 1969), and response priming (Hock & Egeth, 1970; Morton, 1969).

REFERENCES

Dalrymple-Alford, E. C. Interlingual interference in a color-naming task. *Psychonomic Science*, 1968, **10**, 215–216.

Dyer, F. N. The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, 1973, **1**, 106–120.

Egeth, H., Blecker, D. L., & Kamlet, A. S. Verbal interference in a perceptual comparison task. *Perception & Psychophysics*, 1969, **6**, 355–356.

Fox, L. A., Shor, R. E., & Steinman, R. J. Semantic gradients and interference in naming color, spatial direction, and numerosity. *Journal of Experimental Psychology*, 1971, **91**, 59–65.

Hock, H., & Egeth, H. Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology*, 1970, **83**, 299–303.

Klein, G. S. Semantic power measured through the interference of words with color naming. *American Journal of Psychology*, 1964, **77**, 576–588.

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, **76**, 165–178.

Preston, M. S., & Lambert, W. E. Interlingual interference in a bilingual version of the Stroop color-word task. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 295–301.

Pritchatt, D. An investigation into some of the underlying associative verbal processes of the Stroop colour effect. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 351–359.

Stroop, R. J. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, **18**, 643–662.

Thorndike, E. L., & Lorge, I. *The teacher's word-book of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

Treisman, A., & Fearnley, S. The Stroop test: Selective attention to colours and words. *Nature*, 1969, **222**, 437–439.

Warren, R. E. Stimulus encoding and memory. *Journal of Experimental Psychology*, 1972, **94**, 90–100.

(Received September 15, 1973)

Journal of Experimental Psychology
1974, Vol. ..., No. 5, 875–880

E[...] OF FAMILIARITY AND SEQUENCE LENGTH ON ANALOG [...]ATCHES IN THE SIMULTANEOUS MATCHING TASK [1]

GAIL A. BRUDER [2] AND WAYNE SILVERMAN [3]

State University of New York at Buffalo

The study compared the effects of string lengths and familiarity on reaction time (RT) to physical and analog matches in an attempt to determine the nature of a size variation process assumed to occur during analog matches. The results replicated prior findings that "same" RTs to familiar (word) strings were faster than to nonfamiliar strings, and that this advantage increased over string length. For nonfamiliar strings analog matches were slower but this difference was not influenced by string length. Thus, adjusting for size does not seem to occur in the matching or decision stages which appeared to be sequential. However, same RTs to analog and physical matches did not differ when familiar stimuli were used. This suggests that, if size variation is to be postulated as an early processing stage, familiarity effects must also occur quite early.

As has previous research (Eichelman, 1970), the present study used Sternberg's (1969) methodology to investigate the information processing involved in the simultaneous matching task. Posner (1969) concluded that stimuli can be matched on at least two levels, one related to physical attributes, the other dependent on past experience (e.g., the name of the stimulus). Posner and Mitchell (1967) had found *S*s would respond "same" to physically identical stimuli (AA) much faster than to stimuli identical in name only (Aa). In addition, they found matches of stimuli identical in form but not size (Cc), analog matches, to be slightly slower than physical matches, but faster than "name" matches. Posner (1969) indicated that analog matches are based on physical attributes and, in addition, involve a size variation operation.

The purpose of this research was to investigate the effects of familiarity and string length on analog matches in order to see if the size variation process is influenced by past experience (familiarity), and to explore the relationship of size variation and the matching processes. For example, if the size variation process is a global normalizing operation (Neisser, 1966) which occurs prior to any stimulus segmentation or analysis, it should be uneffected by string length or familiarity.

In addition, the study provided additional evidence relative to the effects of familiarity on physical matches. Although Posner and Mitchell (1967) found no familiarity effect on physical matches, more recent work (Eichelman, 1970) has shown that past experience does influence physical matches when letter strings (more than one letter) are used as stimuli. Eichelman (1970) found reaction time (RT) to physical matches was much faster when the letter strings formed words. Because physical matches were still faster than name matches he concluded that the increase in RT of physical matches with words did not reflect reading or naming the words. Eichelman (1970) also found the familiarity effect increased over string length. While maintaining the physical match was based on visual processing, he concluded that familiarity can effect processing stages related to the physical characteristics of the stimuli. In the present study it was hoped that the relationship of familiarity to the size normalizing operation could provide

[1] The research was supported in part by institutional funds (Biomedical Sciences Support Grant from the U. S. Public Health Service) administered by the Research Foundation of the State University of New York. The authors wish to thank Russell Ryden for his assistance in preparing the materials and collecting data, and Webb Stacy and Irv Biederman for their helpful suggestions.

[2] Requests for reprints should be sent to Gail A. Bruder, Psychology Department, 4230 Ridge Lea, SUNY, Buffalo, New York 14226.

[3] Now at the New York State Institute for Basic Research in Mental Retardation.

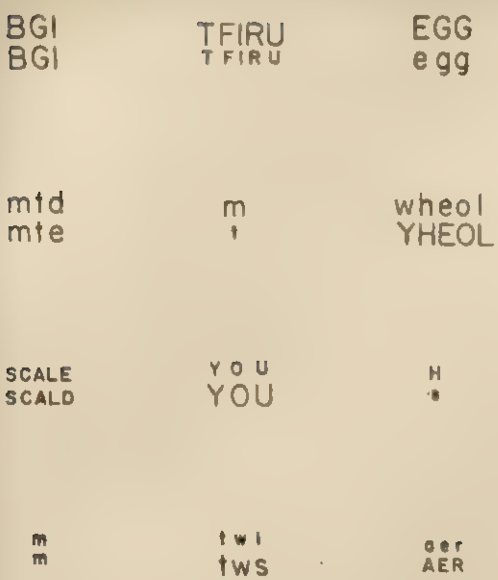


FIGURE 1. Examples of stimuli corresponding to print type and conditions as listed in Table 1.

evidence regarding how familiarity influences physical matches. For example, if the evidence from unfamiliar strings indicates that size normalizing is done separately for each unit (letter) to be matched, the effect of size variation on word matches could indicate if the units to be matched are larger than single letters.

The present study compared physical and analog matches for word and nonword letter strings of 3 and 5 letters. In addition single letter (nonword) matches were used to provide a measure of the effect of the size variation operation with single letter stimuli. In addition, name matches were included to try to replicate Eichelman's (1970) results and thus test his conclusion that physical matches of words do not include reading or naming the words.

METHOD

Design. The study used an incomplete 3 × 3 × 2 × 2 × 2 design. There were three types of matches; physical, analog, and name matches. String lengths of 1, 3, or 5 letters were used. The strings were the same or differed in one letter pair, and the letter strings were either words or nonwords. The design was incomplete because the single letter stimuli were never words. In addition separate scores were obtained for each condition for the first and second half of the test trials.

Subjects. The Ss were 12 male a[illegible] female psychology students. For half the S[illegible]eft response key was assigned to "same" r[illegible] and for the remaining half it was assigned t[illegible]rent" responses. Tray order was also co[illegible]nced across Ss.

Apparatus. The stimuli were proj[illegible] from a Kodak 900 Carousel slide projecto[illegible] h atter mechanism connected to the project[illegible] ontrol circuits controlled the stimulus durat[illegible] inter-stimulus interval. The opening of[illegible] hutter triggered a Hickock timer and the tim[illegible] opped by the depression of one of two r[illegible] keys operated by the index and middle [illegible] S's preferred hand. A Hickock printer r[illegible] d the response and the reaction time. The[illegible] were projected onto a screen 2 m. in front of[illegible]

Materials. The stimuli were lettered [illegible] k on white paper using a Leroy lettering set [illegible] tives of the stimuli were used in slides su[illegible] t the stimuli appeared as white letters on a [illegible] grey background.

There were 10 examples of stimuli for [illegible] of the conditions. Eight stimuli were used fo[illegible] test trials; two were used for the practice tria[illegible] Differences in case were used to obtain nam[illegible] tches. With the lettering set used, upper and l[illegible] case forms for the letters C, O, S, V, W, X, and [illegible] ffered only in size. These letters occasionally o[illegible] d in name match stimuli. Differences in the size of printing were used to create analog match stimuli. All conditions had stimuli printed in each of the possible cases and sizes. The four variations in printing for each of the type of matches are described in Table 1. Examples of the stimuli corresponding to print variations are shown in Figure 1. The visual angles for small lower and upper case letters were about 24′ and 36′, respectively. The visual angles for the large lower and upper case letters were about 36′ and 48′. As can be seen from Figure 1, the visual angles for letter strings depended in part on the size and case of the letters used. The range of

TABLE 1

TYPE OF PRINTING USED FOR THE FOUR STIMULUS EXAMPLES PER CONDITION

Position of spring	Type of match					
	Physical		Analog		Name	
	Size	Case	Size	Case	Size	Case
Example 1						
Top	Large	Upper	Large	Upper	Large	Upper
Bottom	Large	Upper	Small	Upper	Large	Lower
Example 2						
Top	Large	Lower	Large	Lower	Large	Lower
Bottom	Large	Lower	Small	Lower	Large	Upper
Example 3						
Top	Small	Upper	Small	Upper	Small	Upper
Bottom	Small	Upper	Large	Upper	Small	Lower
Example 4						
Top	Small	Lower	Small	Lower	Small	Lower
Bottom	Small	Lower	Large	Lower	Small	Upper

TABLE 2

PERCENT ERRORS PER CONDITION

nulus type	Physical		Analog		Name	
	Same	Different	Same	Different	Same	Different
igle letter	.78	3.91	1.56	5.47	4.69	3.91
etter word	.78	3.91	1.56	6.25	6.25	7.03
letter nonword	3.91	7.03	5.47	10.16	28.12	11.72
letter word	4.69	17.97	3.12	10.94	8.59	11.72
letter nonword	5.47	21.88	4.69	26.56	19.38	19.53

l angles for the three and five letter strings 1°12′ to 2°24′, and 2° to 4°, respectively. The ht of the pairs ranged from 1°12′ to 2°24′ in al angle. There were two stimuli for each iation in printing; one of these appeared in the rst half of the test trials, the other appeared in the second half. Print type for the practice stimuli were chosen in a random fashion.

For the single letter conditions, all letters of the alphabet, except A and I were used. For the different stimuli conditions the letters were randomly paired. For both the word and nonword conditions, common three and five letter words were selected (A or AA words taken from Thorndike & Lorge, 1944). For the different word stimuli condition a igle letter was changed. The letter change was ne such that both strings would be words of approximately equal frequency. The nonword strings were made by randomly arranging the letters of the words to form meaningless letter strings. The different nonword stimuli were obtained by changing a single letter. The position of the different letter was varied over the eight stimulus examples for each condition. All stimuli were randomly assigned to one of the three match conditions.

There were a total of 240 test slides which were divided into 4 trays of 60 slides each. Each tray had two examples of each condition, and the order of tray presentation was completely counterbalanced across *Ss*.

Procedure. The *S* was seated in front of a table on which were two response keys. The *S* was instructed to depress one "key" if the stimulus pairing was "same," and the other if the pairing was "different." A "same" response was defined as indicating the paired letters had the same name. That is, *S* responded "same" when the letter strings spelled the same thing. One half the *Ss* used their index finger to signal "same" and their middle finger to signal "different," while the remainder responded in the reverse manner. Each *S* was instructed as to the nature of the task and was given at least one session on the 60 practice slides. Both speed and accuracy were stressed. Additional practice was available to *S*, and a few so requested. The slides were shown in a dark room, with the slide screen about 2 meters in front of *S*. Each slide remained on the screen until the *S* responded. There was a 1 sec. intertrial interval and several seconds elapsed between each block of 60 trials while *E* changed the slide trays.

RESULTS AND DISCUSSION

Each *S* received four examples of each stimulus condition in each half of the trials. Median RTs were obtained for each *S* based on the correct responses to the four examples, and the analyses of RT were based on this data. The percent errors for each condition are shown in Table 2. It should be noted that in some conditions the median RTs were frequently based on fewer than four correct responses, but the error rates for the same responses to the physical and analog matches were low, and performance in these conditions was the basis for the major conclusions from this study.

Same reaction times. The main comparisons involved the physical and analog matches. One main analysis assessed the effects of size variation, test half, string length, and familiarity. Since there were no words of single letter length, only string lengths three and five were included in this analysis. The summary for this analysis is provided in Table 3.

Practice had no effect on RT within this study. Although RT decreased slightly over halves, the effect was not significant. None of the interactions with test half even approached significance and test half will not be considered further.

As is clear from Figure 2 and Table 4, the RTs increased with string length. In the main analysis (excluding single letter

TABLE 3

Summary of Analysis of Variance of "Same" Reaction Times as a Function of Size Variation, Test Half, String Length, and Familiarity

Source of variance	Mean square for sources[a]	Error term[b]	F	p
Size variation	34677.42	5253.19	6.60	.05
Test half	51855.83	17705.98	2.93	.25
Size × Half	4662.33	10390.86	<1	
String length	1054408.09	48108.10	21.92	.001
Length × Size	13887.15	4825.61	2.88	.25
Length × Half	1.20	20331.48	<1	
Length × Size × Half	58.62	2865.80	<1	
Familiarity	1047232.43	22680.37	46.17	.001
Familiarity × Size	65040.94	13392.32	4.86	.05
Familiarity × Half	9487.98	8759.17	1.08	.25
Familiarity × Size × Half	209.34	10981.77	<1	
Familiarity × Length	166795.67	7043.09	23.68	001
Familiarity × Length × Size	2.34	4008.10	<1	
Familiarity × Length × Half	5752.27	7667.05	<1	
Familiarity × Length × Size × Half	157.03	10129.49	<1	

[a] There is 1 df for all sources of variance.
[b] There are 15 df for all error terms.

stimuli) RTs increased with string length and decreased with familiarity. The interaction of Length × Familiarity reflects the fact that the increase in RT over length was less for words than nonwords. However, a separate analysis indicated the increase in RT over length for words alone was significant, F (1, 15) = 12.15; $p <$ .005. The analyses of nonword stimuli (including single letters) indicated a highly significant and linear increase in RT over length, F (2, 30) = 41.68; $p < .001$. The linear component accounted for 98% of the variance, and the increase was about 76 msec. per letter. The linear trend is consonant with the hypothesis that a sequential letter-by-letter matching process occurs for the nonword stimuli. With only two string lengths, the linear trend for word stimuli could not be assessed. These results support the findings of Eichelman (1970) that familiarity reduces RT and this advantage increases over string length.

Size variation effects were examined to determine the nature of any size normalizing process. According to the main analysis the size differences increased RTs, but the size difference did not interact with string length (analysis excluded single letter stimuli). For just nonword stimuli (all string lengths) the disadvantage of analog compared with physical matches did not

TABLE 4

Mean Reaction Time According to Condition

Pairs	Physical		Analog		Name	
	Word	Nonword	Word	Nonword	Word	Nonword
Different						
Single letter		762		779		763
Three letter	819	834	867	877	901	940
Five letter	952	1,056	897	1,117	969	1,042
Same						
Single letter		652		701		780
Three letter	722	766	728	837	709	1,139
Five letter	814	961	790	1,001	918	1,194

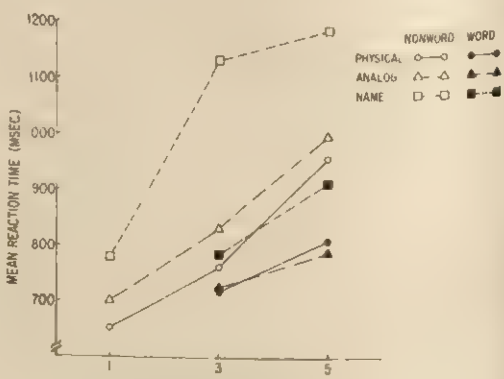


Figure 2. "Same" reaction time as a function of number of letters per string, type of match, and familiarity.

increase over string length, $F < 1$. For word stimuli alone, analog matches were not slower and the interaction, F (1, 15) = 2.91; $p < .25$.

These results provide no support for the hypothesis that size normalizing operations are done sequentially on each letter prior to or during the matching process, which seems to be sequential. However, the idea that the whole stimulus is "normalized" prior to segmentation or analysis is contradicted by the significant interaction obtained between Familiarity and Size Variation. The interaction reflects the fact that size variation only affected performance on nonword stimuli. The analysis for nonword stimuli (including single letter stimuli) showed analog matches to be slower than physical matches, F (1, 15) = 13.11; $p < .005$. There was no difference between analog and physical matches for the word stimuli, $F < 1$.

Different reaction times. The means are presented in Table 3. For all conditions and string lengths the different slides differed in only a single letter.

As Posner and Mitchell (1967) found in their study, RTs to a different single letter pair for all matches were the same and were about equal to the "same" name matches. This is in accord with the conclusion of Posner and Mitchell (1967) that *S*s must check all the way to the name level when a physical difference exists. In addition, the difference in RT between physical and analog matches was less with "different" than with "same" single letter stimuli. This supports the previous conclusion that size normalizing does not occur completely prior to any stimulus analysis.

Because only one type of different condition was used in the present study, the "different" RT data are of limited use. For example comparisons of "same" and "different" RTs with physical or analog matches are difficult to interpret since the different response may include name matching when a physical difference occurs. It was noted earlier that different single letter stimuli are probably processed to the name level in all conditions. However, it is difficult to say if the different pair is always processed to the name level in the letter strings. The position of the different pair may determine the level of processing. For example, if processing is sequential and the previous several letters have been physically identical, a physical difference may be sufficient to initiate a "different" response prior to any name analysis.

Determination of where in processing to place a size normalizing operation was more difficult than expected. The fact that size variation influenced RT to single letters and nonwords but not words precluded placing the operation prior to any stimulus analysis. The fact that size differences did not increase with string length indicates that size normalizing did not occur during the sequential matching of letters. Placing the operation later in a response selection stage is intuitively unappealing and would still leave unexplained the lack of effect of size variation on word stimuli.

Although the physical and name match data in this study replicate Eichelman's (1970) results, the analog match data suggest a complex explanation will be required to explain RT differences in matching word and nonword stimuli. The main question is whether words in a physical match condition are matched on the basis of physical characteristics and/or on some other basis, such as name information. Eichelman (1970) concluded that physical characteristics of words are involved in the matching decision. The faster RT with words could be explained by an increase in the size of the unit to be matched. However, if familiarity only increases the size of the unit to be matched (for example, making a three letter word string equivalent to a single letter) the size variation should have the same effect on words as on nonwords and single letters.

On the other hand, Eichelman (1970) argued quite effectively against the claim that words are being read and matched at the name level. He found name matches to be slower than physical matches; in the present study name matches were also slower than physical matches, for word stimuli, F (1, 30) = 14.86; $p < .005$.

The following discussion will suggest two possible approaches to integrate the effects of physical characteristics and more abstract characteristics related to words into a model of the matching process.

In order to retain a sequential stage model, it could be postulated that a word–nonword decision is made at an early stage. Subsequent processing of words might be done in a different manner (or place) from that of nonwords or letters. (See Biederman, 1972, for a discussion of contingent processing.) Kimura (1967) presented evidence that speech stimuli are processed in the left hemisphere, auditory patterns in the right. More recently, Cohen (1973) found processing differences between hemispheres with visually presented linguistic material. Johnston and McClelland (1973), comparing perception of words and letters, suggested that different systems may exist for processing letters and words. In the contingency model suggested here, physical matching of the whole word would be done with size variation filtered or ignored. If a physical match were not obtained, further processing would be done at the name level. Processing of nonword stimuli would be done letter-by-letter, with failure to obtain a physical match leading to matching at the name level. Filtering of size variation would not apply to nonword stimuli (including single letters).

On the other hand, parallel processing of physical and abstract characteristics of stimuli might be suggested. A redundancy gain between processing physical and abstract characteristics might account for RT differences between words and nonwords (Biederman & Checkosky, 1970). An alternative is an interactive parallel model.[4] In this type of model a same–different decision would be based on a composite of information from physical and abstract comparisons. If processing of abstract characteristics is faster, or takes priority, the "same" decision for physical and analog matches would be made more quickly for words than nonwords, which do not have abstract "word" characteristics. "Same" decisions might be slower for name matches because of interference from physical comparisons indicating that differences exist. One remaining problem is to specify why information about physical *size* differences would not interfere with a "same" response for words.

Clearly the development of an adequate model will require additional experimentation. The results of the present study suggest an adequate model will have to take into account the fact that physical characteristics relating to letter shape influenced word and nonword matches, while the physical characteristic of size influenced only the matching of nonwords and single letters.

[4] This model, based on Thomas (1969), was suggested by Webb Stacy.

REFERENCES

Biederman, I. Human performance in contingent information-processing tasks. *Journal of Experimental Psychology*, 1972, **93**, 219–238.

Biederman, I., & Checkosky, S. F. Processing redundant information. *Journal of Experimental Psychology*, 1970, **83**, 486–490.

Cohen, G. Hemispheric differences in serial versus parallel processing. *Journal of Experimental Psychology*, 1973, **97**, 349–356.

Eichelman, W. H. Familiarity effects in the simultaneous matching task. *Journal of Experimental Psychology*, 1970, **86**, 275–282.

Johnston, J. C., & McClelland, J. L. Visual factors in word perception. *Perception and Psychophysics*, 1973, **14**, 365–370.

Kimura, D. Functional asymmetry of the brain in dichotic listening. *Cortex*, 1967, **3**, 163–178.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1966.

Posner, M. I. Abstraction and the process of recognition. In G. Bower & J. T. Spence (Eds.), *Psychology of learning and motivation*, Vol. 3. New York: Academic Press, 1969.

Posner, M. I., & Mitchell, R. F. Chronometric analyses of classification. *Psychological Review*, 1967, **74**, 392–409.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

Thomas, E. A. Alternative models for information processing: constructing non-parametric tests. *British Journal of Mathematical and Statistical Psychology*, 1969, **22**, 105–113.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

(Received September 20, 1973)

Journal ... ntal Psychology
1974, V... 5, 881–887

NSFER OF ILLUSION DECREMENT AS A FUNCTION OF PERCEIVED SIMILARITY[1]

STANLEY COREN[2]
University of British Columbia, Vancouver, British Columbia, Canada

JOAN S. GIRGUS
City College of the City University of New York

Varieties of the Müller-Lyer illusion were scaled for perceived similarity to the standard Müller-Lyer configuration. Five variants differing in scaled similarity were pretested for illusion magnitude. All *S*s then received practice with the standard configuration during which illusion decrement was obtained. This was followed by a posttest on the initially tested variant. Transfer of decrement was monotonically related to perceived similarity. These data indicate central processing components in the Müller-Lyer illusion.

Since the first systematic studies of visual ·ometric illusions were undertaken by)ppel in 1854–1855, numerous theoretical explanations for these perceptual distortions have been offered. These theoretical treatments may be divided into two general classes: (*a*) theories based directly on the structural properties of the optical and nervous system; and (*b*) theories based on the way visual information is processed in he higher centers.

Structural theories contend that the il- l :sory distortions arise from physiological sources associated with the normal optical . d neural processing of visual inputs. Thus Chiang (1968) suggests that simple degradation of the optical array due to diffraction of the light as it passes through the crystalline lens may produce some intersecting line illusions such as the Müller-Lyer, Zoellner, or Poggendorff. Coren (1969) has demonstrated that this source of distortion can account for some 15% of the Poggendorff illusion with the other 85% of the illusion presumably caused by other factors. Another structural source of illusory distortion has been suggested by von Békésy (1967) and Ganz (1966) who contend that lateral inhibitory interactions operating on converging or intersecting line elements cause some of the perceived contour displacements and distortions. Coren (1970) and Girgus, Coren, and Horowitz (1973) have shown that removal of such converging line elements results in a reduction in the magnitude of several classical illusions, although significant distortions still remain in all cases.

Process theories attribute the existence of illusions to the cognitive processing of the perceptual inputs. Thus, Gregory (1963) and Gillam (1971) have shown that perspective cues which inappropriately evoke constancy scaling mechanisms contribute to the formation of some classical illusions. Coren (1971), Girgus, Coren, and Agdern (1972), Massaro and Anderson (1971), Pressey (1967), and Restle and Merryman (1968) have all found evidence that comparative judgmental components, such as cognitive contrast play a part in illusion formation, especially in configurations with unequal sized elements such as the Ebbinghaus figure. On the other hand, Carr (1935), Coren and Girgus (1972a), and Erlebacher and Sekuler (1969) have provided evidence that confusion of test with inducing elements seems to contribute to illusion magnitude in some configurations. A recently revived theoretical position maintains that some illusory distortions are based on information from scanning eye movements over the figures or information extracted to guide such movements (Binet, 1895; Coren & Hoenig, 1972a; Van Biervliet, 1896).

[1] We gratefully acknowledge the assistance of Richard Fraenkel in the collection and analysis of these data.

[2] Requests for reprints should be sent to Stanley Coren, Department of Psychology, University of British Columbia, Vancouver 8, British Columbia, Canada.

As early as 1902, Judd observed that, with free inspection over a period of time, the magnitude of the perceptual distortions in the Müller-Lyer illusion diminishes. This has since been repeatedly confirmed both for the Müller-Lyer illusion and for a number of other configurations. Although the precise mechanisms for this decrease in illusion magnitude are not fully understood, it seems unlikely that the peripheral structural components, such as optical aberrations from the crystalline lens, could be responsible for such an illusion decrement. These factors would not be expected to show changes which would lead to reduction in illusion magnitude simply as a function of inspection of the stimulus for a few minutes. There are, however, some more central structural components that have been proposed specifically to account for illusion decrement. These include cortical satiation effects (Kohler & Fishback, 1950) and adaptation of specific feature analyzers (Coltheart, 1971).

There is some evidence that these central structural components are probably not responsible for decrement. Köhler and Fishback (1950) imply maximal decrement in illusion configurations where there are converging line segments; yet, Girgus et al. (1973) have shown that the rate of illusion decrement is independent of the number of converging line elements. Both Köhler and Fishback (1950) and Coltheart (1971) predict the maximum reduction in illusion strength under fixation conditions; yet several investigators have shown that free inspection leads to a larger reduction in illusion magnitude than fixation (Coren & Hoenig, 1972b; Festinger, White, & Allyn, 1968).

There is some further indirect evidence that decrement of illusion magnitude involves process rather than structural components: (*a*) the rate of decrement responds to traditional learning variables such as the spacing of trials (Dewar, 1968; Mountjoy, 1958); (*b*) the amount of illusion decrement is cumulative over days and weeks (Judd, 1902); (*c*) direct manipulation of factors which should effect structural mechanisms of decrement such as the number of converging and intersecting line segments, do not appear to alter the rate of illusion diminution with inspection (Girgus et al., 1973).

Although these results suggest that the decrease in illusion magnitude with free inspection is due to process components, they are not conclusive. There is a technique, however, emanating from traditional learning studies, which may provide a tool for answering this question in a more convincing manner. Consider an *O* who has been inspecting an illusion configuration such as the Müller-Lyer, and has been allowed to observe it long enough for a reasonable amount of decrement to take place. If we now present him with a variant of the inspected illusion which differs in some respects from the originally viewed form, it is not unreasonable to expect some transfer of decrement to this new configuration. Some evidence for such transfer of illusion decrement may be inferred from Parker and Newbigging (1963). This transfer, or generalization, would be expected on the basis of both structural and process explanations. However, the parameters which would affect the amount of transfer are different for the two classes of illusion theories. Structural explanations would predict that the amount of transfer of illusion decrement would depend on the number of common elements in the two stimuli. Thus the greater the number of formal configurational similarities, the greater the expected amount of transfer. A process explanation would contend that the amount of decrement transferred to the new variant would depend upon the observers cognitive assessment of the similarity between the configurations. The more similar the figures appear, the greater the amount of transfer of the reduction in illusion magnitude, regardless of the number of elements in common. Thus the amount of transfer or generalization of illusion decrement seems like a useful means of clarifying the theoretical nature of illusion decrement.

METHOD

Scaling of stimuli. Since this experiment is designed to explore whether illusion decrement trans-

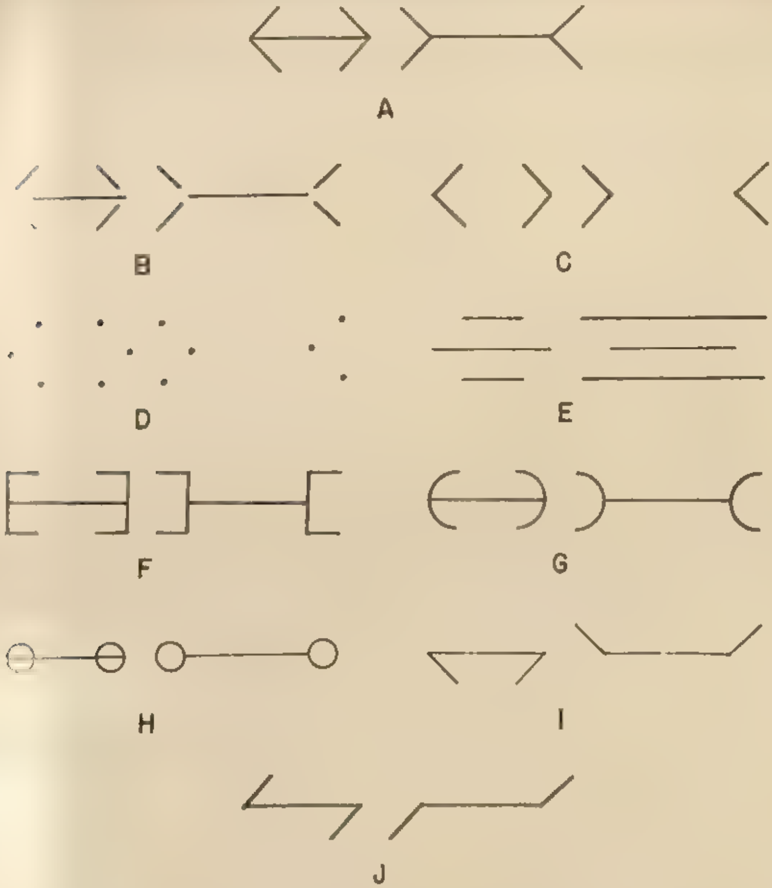


FIGURE 1. Variants of the Müller-Lyer figure scaled for similarity to the standard figure. (Forms A through E are used in the main experiment: [A] Standard form, [B] Exploded form, [C] Shafts removed, [D] Dot form after Coren, [E] Line form after Piaget, [F] Squared wings, [G] Curved wings, [H] Full circles after Parker and Newbigging, [I] Symmetrical, and [J] Asymmetrical incomplete forms after Sanford.)

fers more to perceptually similar configurations than to dissimilar ones, it is first required to establish a scale of similarity among illusion configurations. Ten variants of the Müller-Lyer figure were used to obtain such a metric. They consisted of the standard configuration (Figure 1A), a variant in which the wings are separated from the shaft by a small gap (1B), a standard configuration with the horizontal shafts removed (1C), a dot form of the illusion (1D) which was introduced by Coren (1970), an all horizontal version of the illusion (1E) introduced by Piaget (1969), a figure with squared off wings (1F), one with half circles as wings (1G), another with appended full circles (1H), after Parker and Newbigging (1963), and two configurations from Sanford (1898) that use only half of the arrowheads, either symmetrically (1I) or asymmetrically (1J) oriented. Each configuration was drawn on a 20 × 25 cm. piece of stiff paper with 1-mm. wide black lines. The horizontal shafts (or distance between the shafts when the shafts were not present) were always 8 cm. in length. Each of 15 *S*s was asked to judge the similarity of each configuration, presented one at a time, to the standard configuration (1A) which was continuously visible. Judgments were made using a 7-point rating scale labeled (1) extremely similar, (2) very similar, (3) fairly similar, (4) medium, (5) fairly different, (6) very different, (7) extremely different. The mean of the ratings per stimulus configuration can be used to indicate the perceived similarity between the configuration and the standard version of the illusion. The obtained ratings were used to select four variants of the illusion reliably rated as differing from each other in their similarity to the standard configuration. Those selected were the exploded form (1B) with a mean rating of 1.47, the configuration without shafts (1C) with a mean

rating of 3.47, the dot form (1D) rated 4.60, and the Piaget form (1E) whose rated similarity was a distant 6.53. *T* tests indicated that the mean rating of each of these configurations was significantly different from that of each of the other configurations ($p < .05$). It should be clear that this particular scaling procedure gives only a one-dimensional representation of the psychological space. However, since the theoretical issue involves only the question of the psychological distance between the standard form and the respective variants, this should be sufficient.

Subjects. Fifty undergraduate volunteers served as *S*s. Each was randomly assigned to one of the five variants of the Müller-Lyer illusion used in this study (the standard configuration plus the four variants described above).

Stimuli and apparatus. The five (Müller-Lyer) configurations—1A, 1B, 1C, 1D, 1E—served as stimuli. All of the figures were constructed with 1-mm. wide black lines. The reflectance of the lines was 5.1% and of the background was 74.3%. The test shafts (or the distance between the wings) were 8 cm. long, which corresponded to a visual angle of 12.0. When angles were present, they were 45° from the horizontal; when wings were present, they were 2 cm. long. For each configuration, the apparently longer half of the illusion was made adjustable through the use of a tongue and groove arrangement. Readings were taken in mm. from a scale affixed to the back of the adjustable half of the illusion.

Procedure. Each *S* first made two settings to apparent equality for one of the five variants of the illusion, depending upon his assigned condition. Settings were balanced as to starting position, half starting at clearly longer and half at clearly shorter. Following this, all *S*s then inspected the standard form of the Müller-Lyer figure (1A) for 5 min. with free eye movements. During this period *S*s made settings of apparent equality at 1-min. intervals. Previous investigations have shown that this procedure leads to a sizable reduction in illusion magnitude (Coren & Girgus, 1972b; Girgus et al., 1973). Following the inspection period, all *S*s made another two settings of apparent equality on the initially tested illusion variant. Any diminution in the strength of the nonpracticed illusion configuration between the pretest and the posttest may be taken as an indication of the amount of illusion decrement that has transferred from the standard configuration to that particular variant.

Results and Discussion

Since we are interested in the amount of transfer of illusion decrement from the standard Müller-Lyer configuration to other variants of the illusion, we must first verify that illusion decrement has in fact taken place. An analysis of variance performed on the judgments made during the 5-min. inspection period with the stan rd form of the illusion reveals that the ount of illusion decrement obtained acr the five groups is statistically significant. (5, 225) = 9.96, $p < .001$. However, e is no significant difference among the oups in amount of decrement and no nificant interaction. The mean amoun f decrement was 7.0 mm. This ma itude is quite comparable to values rted in other studies of illusion decre t using the standard configuration (Corer Girgus, 1972a, 1972b; Girgus et al., 1). Regression lines were fitted throug he plot of illusion magnitude against tim or each experimental group. The mea ope of these lines was 1.4 mm/min; range among the five slopes was only .4 m/min and tests on all possible paired d rences yielded no significant group d rences. These results support the conten on that the magnitude of illusion decre ent obtained on the standard figure is equ l across conditions.

The mean amount of illusion per test configuration for the pretest and the posttest is shown in Table 1. In order to assess the amount of transfer from the standard form of the illusion to the five test configurations, an analysis of variance was performed on these data. The results of this analysis indicate that there is a significant pretest–posttest difference, thus demonstrating that transfer of illusion decrement has occurred $F(1, 45) = 73.71$, $p < .01$. In addition a significant interaction between pretest–posttest differences and illusion form indicates a difference among configurations in the amount of transfer, *F*

TABLE 1

Illusion Magnitude of Five Variants of the Müller-Lyer Pre- and Postinspection of the Standard Configuration A

Illusion form[a]	Preinspection illusion (mm.)	Postinspection illusion (mm.)	Transfer (Pre-post)
A	21.9	12.4	9.5
B	15.9	8.9	7.0
C	19.2	15.4	3.8
D	6.6	4.6	2.0
E	8.8	8.5	.3

[a] Forms are labeled as in Figure 1.

(4, 45) = 10.00, $p < .01$. A significant main effect for illusion variants also exists, F (4, 45) = 12.60, $p < .01$. This main effect reflects differences in the initial magnitude of illusion obtained for the various forms, and it is correlated with the number of converging and intersecting line elements in the stimulus as has been reported elsewhere (Girgus et al., 1973). A quick glance at the table also verifies that the magnitude of decrement transferred from the standard to the test variants is independent of the initial magnitude of illusion.

The most theoretically relevant comparison is the amount of transfer as a function of the perceived similarity among the variants. A process theory would predict that as rated similarity between inspection and test figure decreases the amount of decrement transferred should also decrease. The amount of transfer from inspection of the standard is computed from the pretest–posttest differences and plotted against rated similarity of the test configuration to the standard in Figure 2A. It is clear that less transfer is found for more perceptually dissimilar stimuli. A linear trend through these points confirms this, F (1, 45) = 13.04, $p < .01$.

It is important to assess whether these results based upon perceived similarity between the test and the standard configurations predict the amount of illusion decrement transferred with any more accuracy than would a formal analysis of the contour and pattern properties of the configurations. Two methods of assessing the formal similarity of the illusion variants were used to answer this question. In the first, the standard form of the illusion was superimposed on a 40 × 200 matrix of squares, each of which subtended 1 square mm. The standard figure was found to overlap 320 units on this matrix. Each successive variant was then superimposed on the matrix. The number of units held in common between the standard and the variant then serves as an index of the similarity between the patterns. The second scoring procedure simply provides a count of the theoretically relevant pattern elements present in the illusion figures.

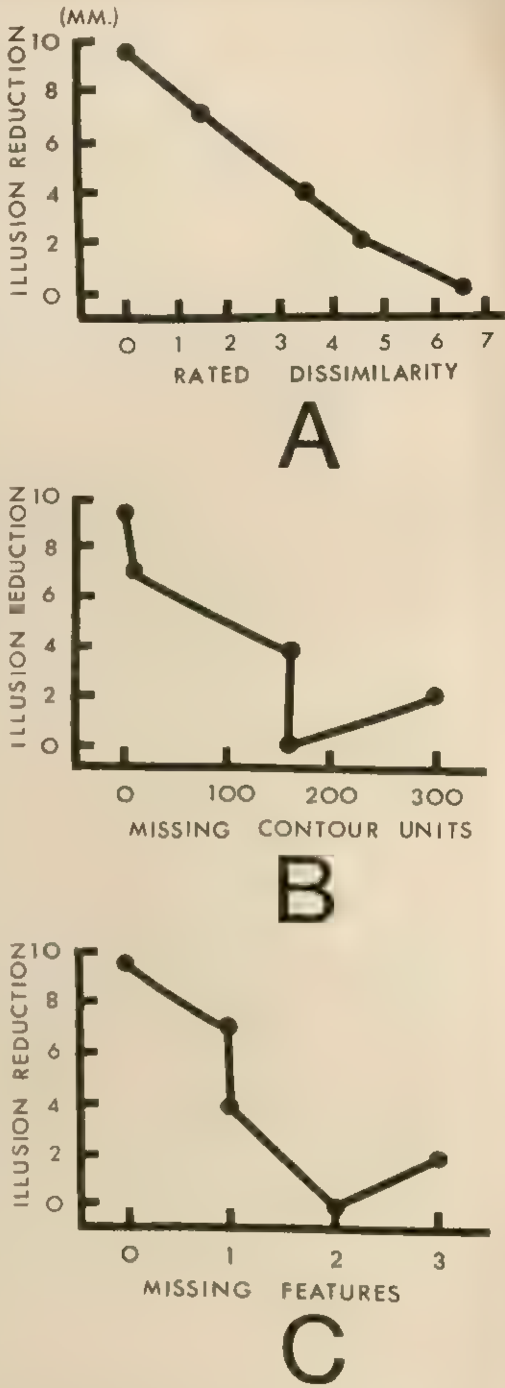


FIGURE 2. Transfer of illusion decrement, measured as illusion reduction in the test figure after practice with the standard form, plotted against (A) perceived dissimilarity of test and standard form, (B) number of missing contour units, and (C) number of missing pattern elements.

These are converging lines, horizontal shaft, and closed angles. All of these properties are contained by the standard form of the illusion, and the number contained by each variant may serve as an index of formal similarity. Figure 2B plots the degree of transfer as a function of the number of missing contour units, and Figure 2C plots the transfer as a function of the number of missing pattern elements. Visual inspection of these figures seems to indicate that neither of the two scoring schemes produces as good a fit to the data as does rated dissimilarity. This is confirmed if we compute the product-moment correlation coefficient for each method of presentation. For rated dissimilarity the correlation with transfer of decrement is $-.751$, for missing contour elements it is $-.448$, and for missing pattern elements it is $-.518$. Although each of these is significantly greater than zero with $p < .01$, the coefficient for the rating procedure is significantly greater than either of the other two procedures with $t\ (47) = 3.47$, $p < .01$, and $t\ (47) = 3.01$, $p < .01$, respectively. This result seems to warrant the interpretation that perceived similarity among configurations is a better index of the amount of illusion decrement transferred from one configuration to another than indices based upon formal properties of the figures involved.

The data presented above seem to be at variance with predictions from structural interpretations of the diminution of illusions with inspection, such as satiation (Köhler & Fishback, 1950) or adaptation of specific feature analyzers (Coltheart, 1971). These theories would predict that transfer would be a function of the formal similarity of the stimuli, such as the number of common elements in inspection and test forms. In the Müller-Lyer figure, the most theoretically important elements are the lines which define the angle between the wings. For satiation theory, this is the region in which the maximum effect is concentrated, and for specific feature adaptation models, these are the stimuli for the feature detectors which are held to fatigue with inspection. These critical angles are clearly present in Figures 1A, 1B, and 1C thus leading to the prediction that the amount of transfer should be equal among these three targets. Contrary to this, we find a decrease in the amount of transfer as the rated similarity decreases. These differences are substantiated by the statistical significance between the adjacent means ($t = 2.98$ and 3.82, respectively, $p < .01$). Since the structural effects should confine themselves to effects pertaining to angles or slanted lines, we would expect no transfer to Forms 1D and 1E since they lack any converging line elements. In addition, they are virtually devoid of any elements in common with the standard configuration. Although the Piaget from (1E), shows no significant amount of transfer, $t\ (252) = .36$, this would be expected both on the basis of missing common elements and the fact that this configuration is the lowest in perceived similarity to the standard configuration of all of the figures tested. We do, however, find significant transfer of decrement for Coren's dot form of the illusion (1D), which also contains no line elements in common with the standard form, $t\ (252) = 2.39$, $p < .01$. Taken together, these data make it unlikely that decrement of the Müller-Lyer illusion under free viewing conditions is due to satiation of cortical fields or adaptation of specific feature analyzers.

Alternatively, it might be suggested that decrement is due to a shift in the way in which the figure is judged. According to this point of view, during the decrement procedure, the O begins to pay more careful attention to the test extent and to ignore the inducing components. Such attentional restriction has been shown to reduce illusion magnitude in the Müller-Lyer (Benussi, 1912; Coren & Girgus, 1972a). If such a restriction of attention were responsible for decrement, however, one might expect that such a set would carry over equally to all alternative forms of the illusion. This would then predict an equal amount of transfer of illusion decrement for all of the configurations, rather than the generalization curve obtained. If, on the other hand, such a judgmental set was only applied as a function of the perceived

similarity of the test to the inspection figure, such a mechanism could explain these results.

The one thing that does seem to clearly emerge from these data is the fact that illusion decrement almost certainly reflects a process mechanism, responding to central cognitive factors, rather than a structural mechanism, responding only to the formal configurational aspects of the stimulus array. This is confirmed by the fact that perceived similarity between practice and test stimuli seems to act as a very accurate predictor of the amount of transfer or generalization of illusion decrement from a practice to a test form. Although this result does not serve to pin down the actual mechanism involved in illusion decrement, it certainly serves to limit the class of explanations that are most likely apt to be valid.

REFERENCES

Benussi, V. Stroboskopische Schienbewegungen und Geometrischoptishe Gestatlttauschungen. *Archiv für die gesamte Psychologie*, 1912, **24**, 169–1

Binet, A. La mesure des illusions visuelles chez les enfants. *Revue Philosophique*, 1895, **40**, 11–25.

Carr, H. A. *An introduction to space perception.* New York: Longmans, Green, 1935.

Chiang, C. A. A new theory to explain geometrical illusions produced by crossing lines. *Perception & Psychophysics*, 1968, **3**, 174–176.

Coltheart, M. Visual feature-analyzers and aftereffects of tilt and curvature. *Psychological Review*, 1971, **78**, 114–121.

Coren, S. The influence of optical aberrations on the magnitude of the Poggendorff illusion. *Perception & Psychophysics*, 1969, **6**, 185–186.

Coren, S. Lateral inhibition and geometric illusions. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 274–278.

Coren, S. A size contrast illusion without physical size difference. *American Journal of Psychology*, 1971, **84**, 565–566.

Coren, S., & Girgus, J. S. Differentiation and decrement in the Mueller-Lyer illusion. *Perception & Psychophysics*, 1972, **12**, 466–470. (a)

Coren, S., & Girgus, J. S. Illusion decrement in intersecting line figures. *Psychonomic Science*, 1972, **26**, 108–110. (b)

Coren, S., & Hoenig, P. The effect of non-target stimuli upon the length of voluntary saccades. *Perceptual and Motor Skills*, 1972, **34**, 499–508. (a)

Coren, S., & Hoenig, P. Eye movements and decrement in the Oppel-Kundt illusion. *Perception & Psychophysics*, 1972, **12**, 224–225. (b)

Dewar, R. E. Distribution of practice and the Mueller-Lyer illusion. *Perception & Psychophysics*, 1968, **3**, 246–248.

Erlebacher, A., & Sekuler, R. Explanation of the Müller-Lyer illusion: Confusion theory examined. *Journal of Experimental Psychology*, 1969, **80**, 462–467.

Festinger, L., White, C. W., & Allyn, M. R. Eye movements and decrement in the Müller-Lyer illusion. *Perception & Psychophysics*, 1968, **3**, 376–382.

Ganz, L. Mechanism of the figural aftereffects. *Psychological Review*, 1966, **73**, 128–150.

Gillam, B. A depth processing theory of the Poggendorff illusion. *Perception & Psychophysics*, 1971, **10**, 211–216.

Girgus, J. S., Coren, S., & Agdern, M. The interrelationship between the Ebbinghaus and Delboeuf illusions. *Journal of Experimental Psychology*, 1972, **95**, 453–455.

Girgus, J. S., Coren, S., & Horowitz, L. Peripheral and central components in variants of the Mueller-Lyer illusion. *Perception & Psychophysics*, 1973, **13**, 157–160.

Gregory, R. L. Distortion of visual space as inappropriate constancy scaling. *Nature*, 1963, **199**, 678–679.

Judd, C. H. Practice and its effects on the perception of illusions. *Psychological Review*, 1902, **9**, 27–39.

Köhler, W., & Fishback, J. The destruction of the Müller-Lyer illusion in repeated trials: I. An examination of two theories. *Journal of Experimental Psychology*, 1950, **40**, 267–281.

Massaro, D. W., & Anderson, N. H. Judgmental model of the Ebbinghaus illusion. *Journal of Experimental Psychology*, 1971, **89**, 147–151.

Mountjoy, P. T. Effects of exposure time and intertrial interval upon decrement to the Müller-Lyer illusion. *Journal of Experimental Psychology*, 1958, **56**, 97–102.

Oppel, J. J. Uber geometrisch-optische Tauschungen. *Jahresbericht der Frankfurter Vereins*, 1854, **55**, 37–47.

Parker, N. I., & Newbigging, P. L. Magnitude and decrement of the Mueller-Lyer illusion as a function of pre-training. *Canadian Journal of Psychology*, 1963, **17**, 134–140.

Piaget, J. *The mechanisms of perception.* (Trans. by G. N. Seagrim) New York: Basic Books, 1969.

Pressey, A. W. A theory of the Mueller-Lyer illusion. *Perceptual and Motor Skills*, 1967, **25**, 569–572.

Restle, F., & Merryman, C. T. An adaptation-level theory account of a relative-size illusion. *Psychonomic Science*, 1968, **12**, 229–230.

Sanford, E. C. *A course in experimental psychology.* Part I. *Sensation and perception.* Boston: Heath, 1898.

Van Biervliet, J. J. Nouvelles mesures des illusions visuelles chez les adultes et les enfants. *Revue Philosophique*, 1896, **41**, 169–181.

von Békésy, G. *Sensory inhibition.* Princeton: Princeton University Press, 1967.

(Received September 24, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 888–889

SHORT REPORTS

RECOGNITION LATENCY FOR A SUBJECTIVELY ORGANIZED LIST[1]

DOUGLAS J. HERRMANN[2] AND JOHN P. McLAUGHLIN

University of Delaware

A previous study showed that after *S*s memorized a list organized into category groupings, recognition latency (RL) to double-word displays (DWDs) was faster to DWDs representing the same category, HORSE–BEAR, than to DWDs representing different categories, LION–RUSSIA. This result was essentially the same when the words in a DWD were both old, one old and one new, or both new. In contrast, the present experiment showed that after learning a list of unrelated words organized into subjective groupings, the representation of groupings in DWDs affected RL only when both words in a DWD were old. Effects of subjective organization on recognition were hypothesized to differ from those of category organization because the semantic bases of subjective organization are too diverse to influence the recognition retrieval process.

Recent research has shown that organization of items during learning affects the latency of subsequent recognition. For instance, McLaughlin and Herrmann (1972) and Herrmann and McLaughlin (1973) have shown that after having memorized words in category groupings, e.g., *animals*, *countries*, *S*s recognized a double-word display (DWD) faster when the words in a DWD represented the same category, e.g., HORSE–BEAR, than when they were from different categories, e.g., LION–RUSSIA. This result was essentially the same for *old* responses (both words in a DWD from the list), *new* responses (both words not from the list), and *mixed* responses (one word in a DWD from the list).

It is often assumed that organizational processes involving relatively unstructured materials such as a list of unrelated words (Tulving, 1968) are the same as for highly structured materials, such as a categorized list. For example, a subjective grouping may be similar to a category grouping in that both list organizations reflect a semantic organization (cf. Schwartz & Humphreys, 1973). Consequently, one might expect subjective groupings to affect recognition latency in the same manner as that observed for category groupings. Alternatively, if subjective groupings are not equivalent to category groupings in organizational characteristics, a different pattern of results might be expected. The purpose of the present research was to investigate these alternatives.

Method. Twelve students participated as *S*s in partial fulfillment of a course requirement.

A three-channel tachistoscope (Scientific Prototype, Model GB) presented a white pre–post exposure field in the first channel, a fixation field in the second channel, and the DWD in the third channel. The two words in the DWD were typed in uppercase and centered one above the other. The response apparatus consisted of three levers on which *S* rested his index, middle, and ring fin[illegible] Latencies were measured to an accuracy of ±[illegible]

Each *S* was run individually in two [illegible] In the first session *S* memorized a list of [illegible] related words, each of which was printed on [illegible] × 5 in. card. Each of three lists, A, B, and C, [illegible] memorized by four *S*s. Each word on a list wa[illegible] member of a different taxonomic category in th[illegible] [illegible] and Montague (1969) category norms, and [illegible] same 18 categories were used across the thr[illegible] Before studying the list, *S* sorted the words in[illegible] groupings which were conceptually similar. At le[illegible] two and no more than seven groupings were allowed (cf. Mandler, 1967). The *S* sorted and then studied the list (for 3 min.) until he recalled all 18 words (within 1 min.) three times in succession and until there was perfect agreement in composition of sorting for three sorts in succession.

The second session, scheduled at *S*'s convenience, began approximately 24 hr. after the first session for 9 *S*s and 48 hr. later for 3 *S*s. Due to the variable delay between sessions, the *S* began the second session by restudying the list. There were no differences in performance as a function of lag between sessions. After restudying and resorting the words, *S* was seated before the tachistoscope and given recognition-test directions as described in Herrmann and McLaughlin (1973).

The recognition test consisted of two blocks of 18 DWDs each. Words for *old* DWDs came from the list which *S* memorized, e.g., List A. For *mixed* DWDs one word came from *S*'s list; the other word in *mixed* DWDs and both words in *new* DWDs came from one of the other lists in Block 1, e.g., List B, and from the remaining list in Block 2, e.g., List C. Thus, all *old* words were presented twice across the blocks, but nonmemorized words were presented only once across blocks. A different random presentation of conditions was used for each *S*.

For half of the *old* DWDs, the two words in a DWD had been studied together in a grouping during acquisition, i.e., the words in a DWD origi-

[1] The authors thank Robert M. Schwartz and Edward J. Shoben for helpful comments on an earlier draft of this article.

[2] Requests for reprints should be sent to Douglas J. Herrmann, who is now at the Department of Psychology, Hamilton College, Clinton, New York 13323.

nated [illegible] the same study grouping (SG-DWDs). For the remaining half of the *old* DWDs, the two words of a DWD originated from different study groupings (DG-DWDs). Similarly, half of the *mixed* and *new* DWDs were SG-DWDs and half were DG-DWDs, wherein grouping origin was designated by the category to which nonmemorized words belonged and by which categories were represented in *S*'s groupings in acquisition. For example, if in memorization *S* had grouped the words WHISKEY and SUGAR, then he was considered to have grouped the categories *alcoholic beverage* and a *substance for flavoring food*. Since it was possible that *S* might organize nonmemorized items differently from memorized items, despite both kinds of items being from the same categories, a comparison was made of the groupings of List A, by the 4 *S*s who learned that list, to the groupings of List B and List C, by the 4 *S*s who learned each of those lists. It was found that there was a significant tendency for the same categories to be represented in a grouping on all three lists, $p < .001$ (see Herrmann, 1972). For example, regardless of which list he learned, *S* tended to place an *alcoholic beverage* and a *substance for flavoring foods* in the same grouping.

Results and discussion. The latency data for correct recognition responses (96.3% of all responses) are shown in Table 1. The table presents the mean of median correct reaction times (RTs) for the three response types as a function of the grouping origin of DWDs, i.e., SG and DG. The medians were based on the three RTs per condition within a presentation block. The latencies in the table have been averaged over blocks, since the analysis of variance indicated that blocks did not significantly affect latency or interact with the other experimental variables, $ps > .05$. Thus, the repetition of *old* words and the lack of repetition of *new* words across blocks did not differentially alter the effect of subjective organization on *old*, *mixed*, and *new* latencies.

Inspection of Table 1 suggests that grouping origin of words in a DWD affected *old* latencies but did not affect *mixed* and *new* latencies. The analysis of variance and simple effects tests confirmed that *old* RT to SG-DWDs was less than RTs for all other kinds of DWDs ($p < .01$), but that all other RTs did not differ significantly from each other. Thus, these data show again that the relationships used during learning affect recognition (cf. Herrmann & McLaughlin, 1973).

The results for subjective organization differ from the effects of category organization on recognition latency. Herrmann and McLaughlin (1973) found latency was faster to same-category DWDs than to different-category DWDs for *new* and *mixed* DWDs as well as for *old* DWDs. In contrast, in the present study, latency was faster to same-grouping DWDs than to different-grouping DWDs for *old* DWDs but not for *new* and *mixed* DWDs. This contrast in results for category groupings and subjective groupings is somewhat surprising since both types of organization have a semantic basis. Apparently, the semantic organization in subjective

TABLE 1

Mean of Median Correct Reaction Times (in msec.) for *old*, *mixed*, and *new* Responses as a Function of Grouping Origin

Response type	Grouping origin	
	Same grouping	Different grouping
Old	1,584	1,797
Mixed	1,945	1,906
New	1,835	1,861

groupings, in the absence of learning, was not sufficient to affect recognition latency. For subjective groupings to affect latency, it seems necessary that the grouping be memorized, i.e., contiguity during learning is crucial for the subjective-organization effect on recognition latency.

Regardless of contiguity's role in the effect on *old* latencies, it is puzzling that the semantic information alone in subjective groupings did not have some effect on mixed and new latencies. Models of semantic memory, e.g., Collins and Quillian (1969), predict that successive retrieval of related items will take less time than retrieval of unrelated items. The lack of an effect from the semantic basis of subjective groupings suggests that subjective groupings differ qualitatively from common taxonomic categories. Subjective groupings in the present study were apparently based on several semantic factors, in contrast to the single categorical basis of category groupings. For example, many *S*s placed *alcoholic beverages*, *substances for flavoring foods*, and *kitchen utensils* in a "foods" grouping. Note that the group included "things to eat" as well as "things to eat with." Thus, a grouping was based on both taxonomic relationships and use relationships. Because of the diverse semantic bases of subjective groupings, semantics may have no influence on recognition retrieval processes when a subjectively organized list is involved. This conclusion does not deny that semantics may influence the formation and initial storage of subjective groupings.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 1969, **80**, (3, Pt. 2).

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240-247.

Herrmann, D. J. The effects of organization in long-term memory on recognition latency. Unpublished doctoral dissertation, University of Delaware, 1972.

Herrmann, D. J., & McLaughlin, J. P. Effects of experimental and preexperimental organization on recognition: Evidence for two storage systems in long-term memory. *Journal of Experimental Psychology*, 1973, **99**, 174-179.

Mandler, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* New York: Academic Press, 1967.

McLaughlin, J. P., & Herrmann, D. J. Recognition of a categorized list. *Journal of Experimental Psychology*, 1972, **96**, 235-236.

Schwartz, R. M., & Humphreys, M. S. Similarity judgments and free recall of unrelated words. *Journal of Experimental Psychology*, 1973, **101**, 10-15.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

(Received May 14, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 890–892

THE EFFECT OF CUE–CRITERION FUNCTION FORM ON SINGLE-CUE LEARNING[1]

L. F. W. DE KLERK[2] AND A. G. VROON

University of Leiden, Leiden, The Netherlands

This investigation is concerned with the effect of the form of the function relating cues to a task criterion upon single-cue learning. Three function forms were used (one in each condition): a J-shaped, a reversed J-shaped, and a U-shaped function form. The results indicated that (*a*) performance gradually improves with training; (*b*) a U-shaped function is more difficult to learn than a reversed J-shaped function, and the latter is more difficult than a J-shaped function. This result is in agreement with the *progression hypothesis*; and (*c*) the relative variances decrease as a function of training. This result does not support the view that single-cue learning is to be conceived of as a two-process theory, which would predict that the relative variance first decreases and then increases.

A single-cue learning experiment requires the S to infer a value of a criterion variable, Y_i $(i = 1 \ldots n)$ from that of a cue variable X_i $(i = 1 \ldots n)$. In fact, S has to learn to utilize the (real valued) function which determines the assignment of Y_i to X_i. The dependent variable for learning is S's prediction of Y_i, say $\hat{Y}_i$, when confronted with X_i.

It has been commonly assumed that S actually processes information in reaching a solution. The underlying assumption is that S is sampling and testing hypotheses in acquiring the function relating criterion to cue values. There is abundant evidence that the universe of hypotheses from which S is sampling is hierarchically structured. Studies in single-cue learning have shown that nonlinear functions are more difficult to learn than linear functions (e.g., Brehmer, 1971; de Klerk & Oppe, 1972). Positive linear functions are easier to learn than negative linear functions (e.g., Björkman, 1965; Brehmer, 1971; Eisler & Spolander, 1970).

According to the theory of single-cue learning, proposed by de Klerk and Oppe (1972), S starts constructing "simple hypotheses" and progresses—if necessary—towards more "complex hypotheses." Their study provided preliminary evidence for this so-called progression hypothesis. The results indicated that when the task required the S to learn a quadratic cue–criterion relation, linear processing preceded quadratic processing. This effect appeared to be highly influenced by the criterion variance.

The present investigation examines the effects of three different quadratic cue–criterion function forms upon single-cue learning. The cue which Ss had to learn to depend upon was related to criterion in either (*a*) a J-shaped manner; (*b*) a reversed J-shaped manner; or (*c*) a U-shaped manner. If the above mentioned results are valid, they will lead to the following predictions. When the cue–criterion function form is a J arc, then it is [illegible] able to assume that S starts constructing a p[illegible] linear function. When a reversed J-shaped [illegible] form is to be learned, S is assumed to sta[illegible] zing a negative linear function. Since it is e[illegible] to learn a positive linear function than a ne[illegible] linear function, it is to be expected that a J-[illegible] d arc is more accessible and easier to learn tha[illegible] eversed J-shaped arc. When the cue–criteri[illegible] function form is U shaped, S is supposed to start [illegible] ucting two linear functions. Therefore, this [illegible] of task function is expected to be more difficult to learn than the other two forms.

Method. The paradigm used in the present study comprised three events that occurred in the learning trials: a stimulus derived from the function cue (X), the S's response $\hat{Y}$, and feedback of the correct value of Y to the subject. The cue values were presented on a scope display. On this display, two marks were present which indicated the endpoints of an imaginary horizontal line. On every trial a dot was presented on this scale. This dot designated the value of X. The values of X could vary from 0 (i.e., the origin of the scale) to 100 (i.e., the right end point of the scale).

Following each presentation of the cue, Ss had to respond by turning a special knob to the estimate ($\hat{Y}$) of the function value, the scale being graded from 0 to 100.

The Ss received feedback after each setting of the knob by presentation of the correct value of the criterion, Y, as a number on a special number display. The value of Y was computed according to $Y = f(X) + e$, in which e is a random component, the value of e being obtained by random sampling from a normal distribution with a mean of 0 and a variance of 16. This variance is referred to as the criterion variance, s_e^2.

Three different experimental conditions were created by using three different functions. These functions were:

$$Y = .015X^2 - .8X + 21, \qquad [1]$$

$$Y = .015X^2 - 2.2X + 91, \qquad [2]$$

[1] This research was supported by a research grant from The Netherlands Organization for the Advancement of Pure Research (Z. W. O.).

[2] Requests for reprints should be sent to L. F. W. de Klerk, Department of Psychology, Stationsplein, 10, 2e Etage, Leiden, The Netherlands.

and

$$Y = .032X^2 - 3.2X + 91. \qquad [3]$$

Th[illegible] criterion function form represented by Eq[illegible] 1 is a J arc, the form that is represented by [illegible]ation 2 is a reversed J arc, and the form of Fun[illegible]on 3 is a U arc.

D[illegible]ing learning, S was sitting at a table in front of the scope display. On this table were mounted the response knob (provided with a scale), the nu[illegible]ber display, and a microswitch. Each learning tr[illegible] started with the presentation of the cue. Then [illegible]s requested to turn the response knob into the [illegible]ed scale position. After this response S could [illegible]t the presentation of a number on the number [illegible]play (i.e., feedback) by depressing the micro[illegible]itch. Twenty such learning trials constituted a [illegible]rning run. The test, which followed upon each [illegible]arning run, consisted of seven different cue values which were equally spaced along the X continuum. During every test, the same seven cue values were presented 10 times without feedback. In every test the cues occurred in different random orders.

Presentation and recording was done automatically by means of punched tape. The output tape contained the specific values of X that were presented during testing and the corresponding values of $\hat{Y}$.

The Ss were told that Y was a function of X plus or minus a random deviation, and that it was their task to discover this function and to use it as accurately as possible during testing. Each S worked under one condition. The Ss made 10 learning runs alternating with 10 test runs. The whole experiment took about 2 hr. per subject. The Ss were paid.

Twenty-four Ss participated in this study. They were randomly divided into three groups with eight Ss in each. The Ss were unfamiliar with this type of experiment.

Results. For each S and for each run the value of D was computed according to $D = \sqrt{\Sigma(Y_i - \hat{Y}_i)^2/n}$, where n is the number of trials.

An analysis of variance was performed on these D scores. The results showed significant main effects of function form, F (2, 23) = 10.86, $p < .001$, and run, F (9, 23) = 25.31, $p < .001$. The interaction between function and run was not significant.

The results of this analysis of variance indicated that Ss' performance improved as a function of learning. In Figure 1, mean D scores are plotted against the number of test runs for each of the three conditions. Performance gradually increased as training progressed and the overall performance was largely dependent upon the function form.

For each S the variance in the responses for each of the seven cues was computed per test. These response variances, s_y^2, were then divided by the criterion variance, s_e^2.

The overall analysis of variance that has been carried out on these relative variances indicated significant main effects of function, F (2, 108) = 132.50, $p < .001$; run, F (9, 108) = 24.20, $p < .001$; and cue, F (6, 108) = 9.09, $p < .001$. Of the interactions only those of Function × Run, F (18, 108) = 6.53, $p < .001$; and Run × Cue, F (54, 108) = 2.05, $p < .01$, were significant.

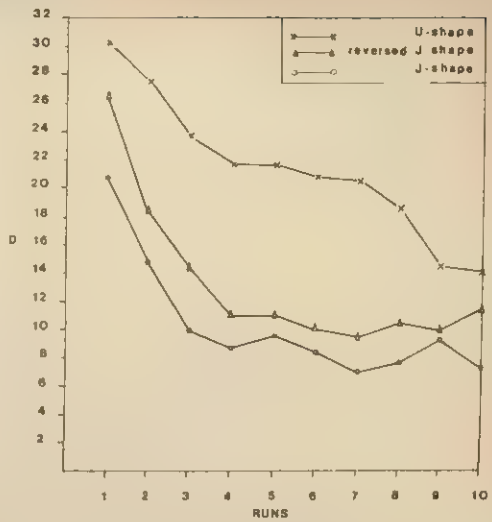


FIGURE 1. Plot of D as a function of test runs for each experimental condition.

The mean relative variances are presented in Table 1. Within every condition the mean variance in Ss' response system exceeded the criterion variance.

Discussion and conclusions. The present results provide several conclusions. First, the data clearly show that performance gradually improved over test runs. The rate of learning was approximately the same for all three functions. The D measures, which reflect the accuracy with which the functions are used, demonstrate that the extent to which Ss learn and utilize relations depends largely upon the task function form. A U-shaped function is more difficult to learn than a reversed J-shaped function; a reversed J-shaped function is more difficult to learn than a J-shaped function. These findings are in agreement with the conjectures that were derived from the progression hypothesis

A second conclusion is that the above mentioned differences between functions remain approximately constant over the range of test runs. Thus, the differences between functions do not decrease as training progresses. This is in agreement with the results of Brehmer (1971), who found that S's skill in using a function depends largely upon the particular function form S has to use. Especially when the task requires S to learn a U-shaped function form, performance appeared to be rather poor.

A third conclusion is that the relative variances, expressed by the s_y^2/s_e^2 ratios, decrease with training. The results indicate that the extent to which these ratios decrease is significantly influenced by the form of the function that is to be learned. Within every condition, the variance in S's response system exceeds the variance in the task system.

TABLE 1

MEAN RELATIVE VARIANCES PER RUN FOR EACH FUNCTION

Function	Run									
	1	2	3	4	5	6	7	8	[illegible]	10
J shaped	10.9	6.2	2.4	1.9	[illegible]	2.2	2.0	1.8	[illegible]	1.7
Reversed J shaped	20.0	11.5	7.5	3.6	5.4	4.5	2.5	3.7	[illegible]	7.1
U shaped	25.6	18.4	16.1	13.1	10.3	10.3	13.5	13.3	[illegible]	9.3

This finding is inconsistent with the results of Bjorkman (1968), who found s_a^2/s_e^2 ratios which were below unity, especially at the end of training.

Bjorkman (1968) suggested that a single-cue probability learning task required the S to learn (a) the function relating criterion to cue values (i.e., functional learning) and (b) the distribution around this function (i.e., cue probability learning). This two-process theory predicts that the relative variance should first decrease (as S learns the function) and then increase (as S learns the distribution of feedback values). Since the present results show only that the relative variances decreases, they do not offer support for this theory. This might be due to two different factors. First, it is possible that 10 successive runs of 20 learning trials only cover the first stage of the learning process and not the second stage (i.e., cue probability learning).

Second, due to S's inability to use a quadratic function correctly, there always will remain a certain amount of (error) variance in S's response system, even when a sufficient number of learning trials is given. The minimum value of s_a^2 [illegible] depends upon the function form that is to [illegible] ned. If this asymptotic value of s_a exceed[illegible] e of s_e, then the relative variance will only [illegible] se as a function of learning. Since in the pr[illegible] dy the value of s_e^2 is relatively small [illegible] derably smaller than in Bjorkman's, 1968, st[illegible] s might be the case here.

REFERENCES

BJORKMAN, M. *Learning of linear functions: [illegible] between a positive and a negative [illegible]* [illegible] Stockholm, Sweden: University of Stockholm, Psychological Laboratories, 1965.

BJORKMAN, M. *The effect of training [illegible] on the response variance in correlational learning.* [illegible] Umea Psychological Rep. 2, Umea, Sweden: Univ[illegible], 1968.

BREHMER, B. Subjects' ability to use fun[illegible] *Psychonomic Science*, 1971, 24, 259–260.

DE KLERK, L. F. W., & OPPE, S. Functional [illegible] under different probabilistic feedback conditions. *Psychonomic Science*, 1972, 29, 377–380.

EISLER, H., & SPOLANDER, K. On the [illegible] in the learning of linear functions. *Scandinavian Journal of Psychology*, 1970, 11, 176–184.

(Received May 21, 1973)

Journal of Experimental Psychology
1974, Vol. 1.., No. 5, 893–896

EFFECTS OF TWO TEMPORAL VARIABLES ON THE LISTENER'S PERCEPTION OF READING RATE

FRANÇOIS GROSJEAN

Département d'Anglo-Américain, Université de Paris VIII, Paris, France

HARLAN LANE [1]

Institut d'Études Linguistiques et Phonétiques, Université de Paris III, Paris, France

Apparent rate grows much more rapidly for the speaker than for the listener as words per minute increase (exponents 2.6 vs. 1.6). The rate of speaking comprises three variables, however: the number of pauses, articulation rate, and the durations of the pauses. When a speaker doubles his rate (as it seems to him) he halves the number of pauses, altering the other variables only slightly. Conversely, the listener is much more impressed by changes in articulation rate than in pausing (exponents 1.5 and −.23). Thus, the perceptual processes of speaker and listener are somewhat complementary, and changes in the two variables have more comparable effects on the listener when expressed in units of apparent magnitude for the speaker. The net effect of mapping autophonic into extraphonic magnitudes in the transmission of speech rate is one of compression, both for overall rate (exponent .6) and for the component articulation rates (.44) and pause frequencies (.21).

A speaker may alter his rate by changing the number of pauses, the articulation rate, or the durations of pauses, in various degrees. The autophonic scale that describes the speaker's perception of his own rate has a slope of 2.6 when expressed as a function of words per minute (in log–log coordinates) and slopes of −1.1, 3.4, and 4.3 when plotted as a function of the respective components of these overall rates (Lane & Grosjean, 1973). It follows that when a speaker varies his rate of reading a known passage, he primarily adds or subtracts pauses (at strategic syntactic locations); he alters articulation rate and pause duration much less. The same is true for spontaneous speech: "Increase of speed in talking is due largely to the closing of gaps and to the heightened continuity with which movements performed at relatively constant rate succeed each other [Goldman-Eisler, 1968, p. 26]." An equivalent way of describing the relative importance of these variables for autophonic rate is to point out that as reading rate increased by a factor of 2.7 (close to the total operational range for reading sentences), the three component variables changed by factors of 10.3, 2.1, and 1.6, respectively.

How the three components of rate control the listener's perception is unknown. We do know from the autophonic study (Lane & Grosjean, 1973), in which a representative speaker's productions were also played back to listeners for numerical estimates, that apparent rate grows much more slowly for the listener than the speaker as words per minute (wpm) increase (slopes of 1.5 vs. 2.6). This difference in slopes indicates that the speaker's perception of his own rate is not based solely on auditory cues. Therefore, pause frequency, articulation rate, and pause duration may not have the same relative importance for the listener's perception of rate as they do for the speaker; it may matter a great deal whether it is your own tongue that is wagging. In the present experiment, then, the two variables that proved most salient in the perception of autophonic rate were decoupled, and each was varied systematically in a 5 × 5 array of rates presented to listeners. We examined the relative contributions of pause frequency and articulation rate to the listener's extraphonic perception of rate, the invariance of the cue hierarchy across

[1] Requests for reprints should be sent to Harlan Lane, Department of Linguistics, University of California, San Diego, La Jolla, California 92037.

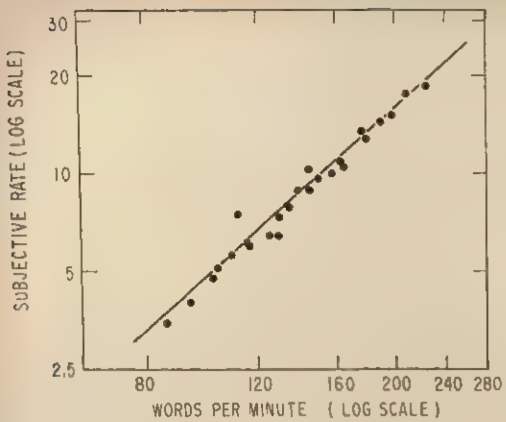


FIGURE 1. The extraphonic scale of reading rate. (Each point is the geometric mean of 68 rate estimates, 4 by each of 17 listeners. The 25 values of words per minute reflect 5 levels of articulation rate at each of 5 levels of pause frequency.)

modalities, and the reproducibility of the extraphonic scale with somewhat different procedures and stimulus ensembles.

METHOD

Seventeen graduate students, native speakers of English with no reported speech or hearing defects, served in groups of three or four, while seated in an audiometric room for 1½ hr. They wore Opelem binaural headsets supplied in parallel by an Opelem 2M-70 studio tape recorder.

The method of magnitude estimation (Stevens, 1956) was used to measure the perception of speech rate. The *E* played a recording of the experimental passage, read with a midvalue of number of pauses (7 pauses or 9.4 syllables/run) and of articulation rate (4.5 syllables/sec). The *E* assigned the numerical value 10 to this standard rate (146 wpm). Next, he played 100 recordings of the passage, and the listener assigned to each a number proportional to its apparent rate. In all, 25 rates of reading by the same speaker, ranging from 85 to 225 wpm, were presented four times, in irregular order. Each rate corresponded to a combination of a level of articulation rate and a level of number of pauses, shown on the coordinates of Figure 2. (The length of the pauses was held constant, with tape splicing, at a mean value of .56 ± .01 sec. Each articulation rate was constant within 1% over the five levels of pause frequency.) The same passage employed in the extraphonic study and comparable ranges of the variables measured there were adopted in the present study. However, the rates presented here are somewhat lower (average wpm 146 vs. 165), thus higher pause frequencies (9.6 vs. 6.3) and lower articulation rates (4.5 vs. 4.8 syllables/sec) are naturally encountered.

The experimental passage contained the following 51 words of text (including contractions), comprising 75 syllables (the number at each pause emplacement shows the lowest pause frequency that entailed pausing there; all higher pause frequencies also used that emplacement):

> as far /24/ as I know /7/ I'm a fairly /24/ normal /13/ fifteen year old /3/ neither a complete /13 psychological /24/ case /7/ nor a cut /13/ above /24/ the others /1/ I listen /24 [illegible] Radio /24/ Luxembourg /7/ my hair /24/ falls forward /13/ in the fashionable /24/ style /3 and I wear /13/ polo neck /24/ sweaters /7 but I don't /24/ consider myself /13/ a great /24/ pop fan.

The hierarchy of pause emplacements is that observed in the autophonic experiment of Lane and Grosjean (1973).

RESULTS AND DISCUSSION

The extraphonic scale of reading rate is shown in Figure 1. The magnitude estimations are well fit by a straight line with a slope of 1.7, in logarithmic coordinates. (The mean and *SD* of the 17 individual slopes are 1.7 and .56., respectively.) The scale is substantially the same as that obtained with five rates of reading in which pausing and articulation rate were normally coupled, not varied independently (slope = 1.5, *SD* = .37; Lane & Grosjean, 1973). A slope of 1.6 is found in the latter study over the range of words per minute employed here; moreover, moderate levels of the two variables in the present experiment (3 × 3 subensemble) yield a slope of 1.62. We conclude that the extraphonic scale of reading rate is reasonably invariant over these differences in stimulus array and procedure, with a slope (exponent in linear coordinates) of 1.6.

The contributions of pause frequency and articulation rate to perceived reading rate may be examined in three different ways, all of which give the same general picture. The perceived-rate solid in Figure 2 suggests that the two components of overall rate contribute about equally over the ranges employed (approximately those encountered by Lane & Grosjean, 1973, in actual productions). An analysis of variance of rate estimates confirms that pause frequency and articulation rate are equally important: the two mean squares are 3,255

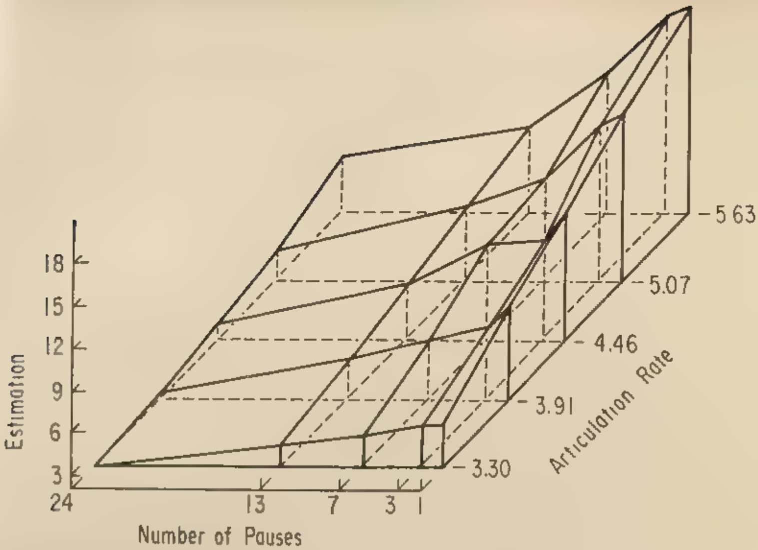


FIGURE 2. Numerical estimates of apparent rate as a function of two component variables, pause frequency and articulation rate. (The 25 line intersections plot the geometric means of 68 rate estimates, 4 by each of 17 listeners.)

and 2,919, respectively; $F(4, 64) = 98$ and 109, $p < .01$. The between-*S*s mean square is 136, *ns*. The interaction of pause frequency and articulation rate, evident in the graph, is statistically significant, mean square = 136; $F(16, 256) = 23$, $p < .01$. Eighty-four percent of the interaction sum of squares is associated with the one *df* corresponding to the bilinear component.[2] This suggests that the two cues combine multiplicatively to yield the net impression of overall rate. Thus, an increase in articulation rate may amplify the effects of pause frequency on apparent rate. The same result would be obtained, however, if the listener combines the two cues linearly under conditions in which the estimates are not linearly related to apparent rate, or if the weights assigned to the cues for integration vary as a function of stimulus magnitude.

A third way to evaluate the relative contributions of pause frequency and articulation rate, and to compare their importance for the speaker and the listener, is to examine the rates of growth of autophonic and extraphonic perceived rate as a function of the two components and overall words per minute. A speaker instructed to double his apparent rate will change pause frequency by a factor of 2, articulation rate by 1.2, and overall rate by 1.3 (since the autophonic slopes are −1.1, 3.4, and 2.6). A listener would estimate, however, that the changes were, in fact, 1.2, 1.3, and 1.5 to 1, respectively (since the extraphonic slopes found in this study are −.23, 1.5, 1.6). Consequently, equal changes in pause frequency and articulation rate have more comparable effects on the listener when those changes are measured in psychological units along the autophonic scale. Extraphonic rate grows approximately six times faster as a function of articulation rate than pause frequency, expressed in physical units. However, the rates of growth as a function of the corresponding apparent (autophonic) magnitudes are more similar: .44 and .21, respectively (absolute values). Thus, the perceptual processes of speaker and listener are somewhat complementary. The net effect is one of compression: extraphonic rate grows as the $1.6/2.6 = .6$ power of auto-

[2] The authors are grateful to Norman H. Anderson for calling this aspect of the results to our attention and for aiding us in applying his theory of functional measurement (Anderson, 1973).

phonic rate. It follows that speakers and listeners will not agree on what constitutes a doubling of rate (or halving, tripling, etc.), and that speakers will always fall short when trying to imitate the rate changes of a model.

REFERENCES

Anderson, N. Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception.* New York: Academic Press, 1973.

Goldman-Eisler, F. *Psycholinguistics: Experiments in spontaneous speech.* London: Academic Press, 1968.

Lane, H., & Grosjean, F. Perception of reading rate by speakers and listeners. *Journal of Experimental Psychology*, 1973, **97**, 141–147.

Stevens, S. S. The direct estimation of sensory magnitudes—Loudness. *American Journal of Psychology*, 1956, **69**, 1–25.

(Received May 31, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 896–898

AROUSAL AND RECALL: EFFECTS OF NOISE ON TWO RETRIEVAL STRATEGIES[1]

STEVEN SCHWARTZ[2]

Northern Illinois University

In a test of the notion that arousal affects the retrieval strategy of pigeonholing by decreasing the separation between risky and cautious criteria, 75 *Ss* heard short meaningful stories accompanied by either No, Medium, or High noise and were asked to recall the characters' names. Common names were better recalled than rare names in each noise condition. Using a decision theory analysis it was found that, in the No Noise condition, *Ss* employed a risky criterion for common names and a cautious criterion for rare names. In noise, however, *Ss* employed a similar criterion for recall of both common and rare names, but sensitivity increased for common names. The results were interpreted as supporting the hypothesis that arousal affects the accessibility of information for retrieval. A possible mechanism for arousal's action, as well as the theoretical implications of these findings, was discussed.

Drawing on his work in perception, Broadbent (1971) has identified two memory retrieval strategies—filtering and pigeonholing. According to Broadbent, filtering occurs when one adopts a "stimulus set," choosing items to be recalled on the basis of some common feature (e.g., acoustic similarity) and ignoring those items in memory without this feature. Filtering, therefore, involves the grouping of input on the basis of physical characteristics. Pigeonholing, on the other hand, occurs when one adopts a "response set" selecting from a large number of items (e.g., a list of words), those constituting a subvocabulary (e.g., the names of colors). Thus, filtering leads to stimulus selectivity, whereas pigeonholing results in response selection. When an operating characteristic is derived relating the probability of a correct response to the probability of an incorrect response (false alarm), pigeonholing produces an increase in the number of correct responses as the false-alarm rate increases, whereas filtering produces a change in the number of correct responses with a constant false-alarm rate. In decision-theory terms, filtering affects d', the difference between the correct response rate and the false-alarm rate, whereas pigeonholing affects β, the critical value of likelihood ratio above which S responds positively.

Despite recent interest in applying decision-theory measures to memory, the effects of arousal on d' and β as estimated from recall data have not been explored. On the basis of vigilance research, however, there is reason to believe that arousal affects pigeonholing but not filtering. Broadbent and Gregory (1965) found that, in a vigilance situation, arousal results in a change in the number of responses assigned to intermediate confidence levels. That is, aroused *Ss* were more certain that a signal was presented or more certain that no signal was presented than unaroused *Ss*. Such a mechanism may produce contradictory effects on pigeonholing.

---

[1] The author is indebted to D. E. Broadbent for his constructive criticism and to a number of students and colleagues whose comments greatly improved this article. Any errors in the final product are, of course, the responsibility of the author.

[2] Requests for reprints should be sent to Steven Schwartz, Department of Psychology, Northern Illinois University, De Kalb, Illinois 60115.

Specifica hen conditions favor a cautious criterion (, ugh) a change in the number of *not sure* respons nay improve performance, whereas, when conditio favor risky responding, arousal may impede rmance.

If al's effect on recall is substantially similar to ct on vigilance, contradictory results would be ted depending on the value of β at the outset order to assess this hypothesis, it is necessa know in advance when *S*s will adopt a risky cr on and when they will adopt a conservative on o that the discrepant effects of arousal may be p ted out. Fortunately, such a situation is e established. The present study employed a m dification of a technique devised by Ingleby 9). In that study, *S*s were asked to recall the names of characters in short stories. Probe mes were suggested, and *S*s were asked to indicate he name was correct and to give a rating of confence in their judgment. Although d' was actually eater for rare names, recall for common names was superior because *S*s employed a riskier criterion (β was lower) for common names. Given a similar experimental task, the major hypothesis of the present study was that arousal would affect pigeonholing by producing a decrease in the separation of criteria (β) for rare and common names while d' remained unchanged. It was hypothesized that β for rare names would decrease, thus reducing the advantage of common names.

Method. The *S*s were 75 introductory psychology udents who received course credit for participaon. All *S*s reported having normal hearing; *S*s rticipated in groups of five. Experimental ses ns were conducted in the university language laboratory. The room was furnished with desks separated by side panels. Each desk contained a set of monophonic earphones.

Three short meaningful stories were used in the present study. The stories, which ranged in length from 375-525 words, were chosen from past issues of the *New Yorker* magazine. Each story had four characters, two of whose names were judged common and two of whose names were judged rare based on the frequency of their occurrence in the Chicago, Illinois, telephone directory. The mean frequency of common names was 920 (of about 900,000 telephone subscribers); the mean frequency for rare names was 43. The stories were recorded and presented on tape.

Upon entering the language laboratory, each *S* was seated at a booth, asked to put on his earphones, and given an instruction and answer booklet. The instructions informed *S* that the experiment was concerned with memory under various conditions. All *S*s were told that they would hear brief stories over their earphones and that a transcript of each story appeared in the answer booklet. Blanks in in the transcripts corresponded to character names. Corresponding to each blank was a suggested name and a rating scale. During the initial story presentation, the room was kept completely dark in order to discourage *S*s from making entries in their answer booklets. At the conclusion of each story, the lights were turned back on, and *S*s were required to read the transcripts and decide whether the name suggested was correct or incorrect and to give a confidence rating in their judgment on a 5-point scale ranging from 1 (certain this is the correct name) to 5 (certain this is not the correct name). For each story, two of the suggested names (one common and one rare name) were correct, while the remaining common and rare suggested names were incorrect. Each *S*, then, had three occasions to judge each of four types of suggested names: a rare name that was correct, a rare name that was incorrect, a common name that was correct, and a common name that was incorrect.

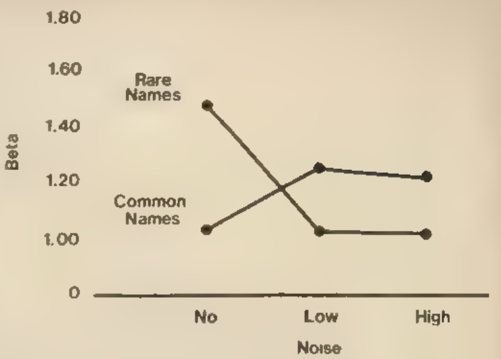


FIGURE 1. Beta as a function of name type and noise. (Each point is the mean β for the five groups in each condition.)

The *S*s were randomly assigned to one of three noise conditions (No, Medium, High). The No Noise group heard the stories on tape, recorded by a male graduate student and broadcast at about 90 db. (A).[3] (The abbreviation db. (A) refers to readings made from the A scale of a sound-level meter, where high frequencies are weighted more than low frequencies.) The Medium Noise group heard the stories accompanied by a continuous white noise, broadcast at about 65 db. (A). The High Noise group heard the stories and 85-db. (A) continuous white noise. For all *S*s, the earphones were silent during the period from the end of one story to the beginning of the next.

Results. Due to the small number of judgments each *S* was asked to make (six for each name type), there was little evidence from false positives for some *S*s in the "certain" category. For this reason, each session's data were pooled across *S*s. Since *S*s participated in small groups of five per session, each noise condition required five sessions. For statistical analyses, then, the unit was sessions and each noise condition had a sample of five.

Following the procedures described by Broadbent (1971, pp. 70–76), β for rare and common names was calculated for all rating categories combined. As Figure 1 illustrates, with increasing noise, β decreased for rare names, but increased for common

[3] Differences in the resistance of the earphones made the sound level estimate only approximate. The actual level for an individual *S* may have varied by about 3%.

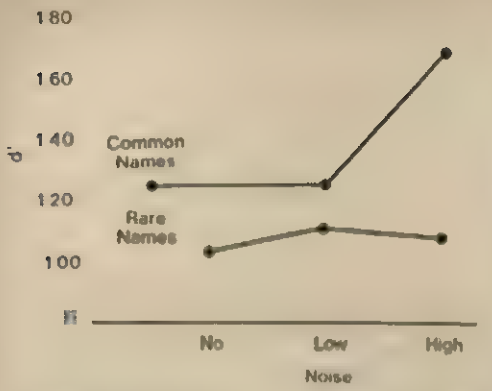


FIGURE 2. Plot of d' at most risky criterion as a function of noise and name type. (Each point is the mean for the five groups in that condition.)

names. An analysis of variance performed on these data confirmed the significant Noise × Name interaction, $F(2, 12) = 4.50$, $p < .05$. An a posteriori test on the mean β scores (Winer, 1962, pp. 309-313) indicated that although β was lower for common names in the No Noise condition ($p < .05$), β did not differ for rare and common names in the two noise conditions.

Despite the decrease in the separation of βs with noise, performance measured by hit rates did not deteriorate for common names relative to rare names. An analysis of variance of hit rates in each confidence level indicated that the hit rate for common names was higher than for rare names at each noise level, $F(1, 12) = 7.71$, $p < .05$.

Since β for rare names did not differ significantly from β for common names in noise, it seems plausible that changes in d' may have been responsible for the continued superiority of common names. Changes in d' are depicted in Figure 2. As indicated, the mean values for d' were 1.03 and 1.24 for rare and common names, respectively, in the No Noise condition, but 1.08 and 1.69 in the High Noise condition. Only the latter difference was significant, $t(4) = 3.0$, $p < .05$.

Discussion. The results of the present study were in accord with those of Ingleby (1969). In the No Noise condition, the condition most directly comparable, common names were recalled with greater accuracy than rare names, despite the fact that d' did not differ for the two types of names. The advantage of common names was due to the lower value of β for these names. The introduction of white-noise-induced arousal led to an increase in β (and pigeonholing) for common names and a decrease in β for rare names. This result supports the hypothesis that arousal improves pigeonholing when conditions favor cautious responding, but [illegible] favored.

The change in pigeonholing [illegible] in the present study, however, did not [illegible] to the predicted changes in overall responding. In [illegible] arousal condition, common names were [illegible] with greater accuracy than rare names. The superiority of common names in noise was due to [illegible] increase in d', which tended to counteract the [illegible] verse effect of noise on β. Thus, it would [illegible] that, for common names at least, arousal [illegible] filtering as well as pigeonholing.

The apparent increase in filtering with increased arousal noted for common names is in line with the results of a study by Hörmann and Osterkamp (1966). They reported a decrease in "cluster[illegible]" in free recall with noise, although similar n[illegible] of words were recalled in the quiet and noise conditions. To the extent that clustering depend[illegible] pigeonholing, noise-induced arousal operated [illegible] manner similar to the findings of the present [illegible]. That is, arousal seemed to make filtering [illegible] important than pigeonholing as a retrieval str[illegible]. Arousal, therefore, will only lead to improved [illegible] if the material may also be recalled by filt[illegible]. The actual effect of arousal on recall would be [illegible] result of the interaction of a number of facto[illegible] cluding the nature of the material to be recalled, [illegible] criterion adopted, and the efficacy of filterin[illegible].

It should be noted that the failure to find [illegible] increase in filtering for rare names in the p[illegible] study may have been due to inherent differe[illegible] discriminability. Thus, although the signal-to-n[illegible] ratio was quite good, it is possible given the [illegible] dynamic range of speech that some loss of aco[illegible] information occurred due to masking. Such [illegible] ing may have affected common and rare [illegible] differentially because common names p[illegible] phonemic characteristics resistant to mask[illegible] because such names contained common lette[illegible] features which were likely to be selectively [illegible] amined by an aroused *S* (cf. Easterbrook, 19[illegible]). In any event, an explanation based solely on [illegible] ing would be hard pressed to account for the c[illegible] in pigeonholing, since β decreased for com[illegible] names, whereas β increased for rare names.

REFERENCES

Broadbent, D. E. *Decision and stress.* New York: Acade[illegible] Press, 1971.

[illegible], M. Effects of noise [illegible] 7, 155-163.

[illegible] emotion on cue utiliza[illegible] tion of behavior. *Psychological* [illegible]

Hörmann, H., & Osterkamp, U. Über den K[illegible] kontinuierlichem Lärm auf die Organization von G[illegible] chtnisinhalten. *Zeitschrift für angewandte Psychologie,* 13, 31-38.

[illegible]. Decision-making processes in human [illegible] memory. Unpublished doctoral dissert[illegible] Cambridge, 1969.

Winer, B. J. *Statistical principles in experimental* [illegible] New York: McGraw-Hill, 1962.

Received June 6, 1973

VERBAL DISCRIMINATION LEARNING FOR BILINGUAL LISTS[1]

ROBERT M. YADRICK AND DONALD H. KAUSLER[2]

University of Missouri

The standard decrement in verbal discrimination learning for an interpair relationship between right (R) and wrong (W) items was found when the related words were bilingual translations of one another. However, the decrement occurred only when the R items were from the primary language (English) and the W items from the secondary language (German). The results were interpreted in terms of the processing tagging of different sets of features for primary language words as [illegible] items than for secondary language words as R items.

[illegible]veral studies (e.g., Eberlein & Raskin, 1968, [illegible]erson & Kausler, 1969, Zimmerman, Shaughn[illegible]., & Underwood, 1972) have demonstrated a [illegible]rement in verbal discrimination (VD) learning [illegible]ciency for lists containing interpair associative [illegible] items. The feature-tagging hypothesis of VD learning (Kausler, 1974, in press) attributes the decrement to the sharing of tagged [illegible] frequency of response cues) [illegible] the related R and W items. False recognitions of W items as being correct presumably occur when the semantic features of these items were [illegible] elsewhere in the list via the rehearsal and [illegible] of their related R items. By contrast, the same [illegible]dies failed to find a comparable decrement for [illegible]s containing intrapair associative relatedness [illegible] [illegible]mpensatory process operates in the intrapair [illegible] [illegible]ing on those features of R items that are [illegible] [illegible] graphic and phonemic features, Kausler, in press [illegible] The previously cited studies employed standard word associates as the means of manipulat[illegible] semantic relatedness. The present study tested the generality of the disparity between interpair and intrapair conditions by employing a different form of semantic relatedness, viz., [illegible] interpair and intrapair experimental [illegible] related R and W items were English-German translations of one another. For the control conditions the bilingual items were semantically [illegible] The *S*s in each condition were students nearing the completion of second year college German.

A question of further interest concerned the differential effects on VD performance of [illegible] secondary language words as R items [illegible] tance of the language variable was suggested by the [illegible] [illegible]t with Spanish-English bilingual *S*s (i.e., Spanish was their primary language, English their [illegible]

formulated within the context of a $W_1 - R_1, W_1 - R_1'$ transfer paradigm (i.e., the R items of List 2 were related to those of List 1, whereas the W items of the two lists were unrelated). Significant mediated facilitation was found, relative to a $W_1 - R_1, W_2 - R_2$ control condition, when the R_1 items were [illegible] words and the R_1' items were t[illegible] secondary language translations, but not when the interlist relationship was reversed. A [illegible] effect in the present study would in[illegible]

[illegible]

The *S*s were 80 students in fourth semester German classes at the University of Missouri. They were paid for their participation in the study, and all of [illegible]

[illegible]

with English words as R items, and 6 pairs with German words as R items) across all conditions prevented the application of the simple rule of always selecting words of one language as R items. For the intrapair condition the R–W pairings were, of course, cross-language translations of one another. For the interpair condition the R–W pairings were formed by re-pairing randomly the English–German translations within the separate sets of the intrapair condition. For the control condition either the English words or the German words of the experimental conditions were replaced by other members of the total word pool. Finally, interitem relatedness, both semantic and orthographic, other than that of direct experimental interest, was minimized as much as possible in terms of *E*'s judgment.

Each *S* received 10 trials by the anticipation method, with the first being a guessing trial. The rate of exposure was 3:3 sec., and the intertrial interval was 6 sec. Standard VD control procedures were employed. That is, the pairs were arranged in different serial orders across trials, the R items appeared equally often on the spatial right and left, etc. The experimental list was preceded by 3 trials on a short practice list conforming to *S*'s specific list condition.

Results and discussion. Means and standard deviations, respectively, for total errors summated across trials and language sets were 18.60 and 9.73, 15.65 and 6.58, and 13.10 and 5.13 for the interpair, intrapair, and control conditions, respectively. Errors were analyzed separately for the two within-list sets. Considering the English set first, the means and standard deviations, respectively, were 11.45 and 8.15, 8.10 and 4.70, and 6.40 and 2.93 for the interlist, intralist, and control conditions, respectively. Comparable values for the German set were 7.15 and 3.03, 7.55 and 3.28, and 6.70 and 3.25.

A 3 × 2 mixed analysis of variance revealed a nonsignificant main effect for list condition, F (2, 57) = 2.77, $p < .10$, a significant main effect for language set, F (1, 57) = 8.94, $p < .01$, and a significant List Condition × Language Set interaction effect, F (2, 57) = 7.76, $p < .005$. In view of the significant interaction effect, the effect of list condition was determined separately for the two language sets. For the English set the between-condition effect was significant, F (2, 57) = 4.13, $p < .02$. A Duncan test yielded a significant comparison between the interpair and control condition means ($p < .01$). Neither the interpair–intrapair comparison nor the intrapair–control comparison reached statistical significance ($ps < .10$ and $> .10$, respectively). Given the pronounced [illegible]rogeneity of variance apparent between list [illegible]ions, the between-condition effect was also analy[illegible] by means of the median test. The between-co[illegible]ion effect was again significant, χ^2 (2) = 7.93, [illegible] .02, but only the contrast between the me[illegible]s of the interpair and control conditions reache[illegible]nificance, χ^2 (1) = 4.95, $p < .05$. Finally, for [illegible] German set the between-condition effect fell [illegible] short of statistical significance, F (2, 57) < 1.

Thus, in agreement with previous [illegible]dies, the present results indicated a decrement [illegible]D learning for a list containing an interpair [illegible]tionship but not an intrapair relationship. H[illegible]ver, the generalization to bilingual relatedness [illegible]estricted by the fact that the decrement occurr[illegible]hen the R and W items were primary and sec[illegible]ary language words, respectively, but not vice [illegible]sa. One possible explanation for the bilingual [illegible]mmetry is that different sets of features are th[illegible] recipients of frequency of response cues wh[illegible] primary and secondary language words function [illegible] R items. Semantic features of primary word item[illegible]re likely to be processed and tagged via thei[illegible] rehearsal during feedback intervals. Since these [illegible]ures are shared by their secondary language co[illegible]terparts, false recognitions of the latter are po[illegible]ble when they are encountered (as W items) du[illegible]ing later anticipation intervals. On the other hand, [illegible]honemic and orthographic features of the less fa[illegible]liar secondary language words are likely to be processed and tagged when they serve as R items. Since these features are not shared by their primary language counterparts, false recognitions of the latter are unlikely to take place when they are encountered during later anticipation intervals.

REFERENCES

Eberlein, E., & Raskin, D. Intrapair and interpair associations in verbal discrimination learning. *Psychonomic Science*, 1968, **11**, 145–146.

Fulkerson, F. E., & Kausler, D. H. Effects of intrapair and interpair bidirectional associates on verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 307–310.

Kausler, D. H. Homographs as double function items in verbal discrimination learning. *Journal of Experimental Psychology*, 1973, **98**, 214–216.

Kausler, D. H. Continuity of processes across variants of recognition learning. In R. L. Solso (Ed.), *Theories of cognitive psychology: The Loyola symposium*. Potomac, Md.: Lawrence Erlbaum Associates, in press.

Kausler, D. H., & Kollasch, S. F. Bilingual verbal discrimination transfer: Implications for verbal discrimination theory. Paper presented at the annual meeting of the Psychonomic Society, San Antonio, November 1970.

Zimmerman, J., Shaughnessy, J. J., & Underwood, B. J. The role of associations in verbal discrimination learning. *American Journal of Psychology*, 1972, **85**, 499–518.

(Received June 18, 1973)

Journal o' ...tal Psychology
1974, V... 5, 901–903

.TITION EFFECTS IN ICONIC AND VERBAL SHORT-TERM MEMORY[1]

DEREK BESNER,[2] J. K. KEATING, LESLIE J. CAKE, AND RICHARD MADDIGAN
Memorial University of Newfoundland, St. John's, Newfoundland, Canada

The Sperling 1960 procedure of full and partial report was used with tachistoscopically presented arrays of 100-msec. duration in two separate experiments. A digit slide was repeated 48 times per session, with 48 slides of nonrepeated series (digits in Experiment I; letters in Experiment II) interpolated between repetitions. The results of both experiments demonstrated better recall of the repeated array. It was concluded that Turvey's 1967 observation that repetition effects differentiate verbal and iconic memory may be incorrect.

...perling (1960) viewed the difference between ...l and full report of tachistoscopic arrays as ...lecting qualitative as well as quantitative memory ...fects. That is, recall under full report was inter...reted as being from a verbal short-term store, ...hile recall under partial report was assumed to ...flect a sensory (iconic) trace. Given the existence of a separate iconic memory store, an important question concerns the kind of operations that might serve to illustrate different properties of these two memory systems.

In support of this notion Turvey (1967) reported two experiments which indicated that partial report recall of a repeated 3 × 5 digit matrix was no ...different from that of nonrepeated matrices interpolated between the successive repetitions. These ...sults were considered to be consistent with the ...ea that a sensory trace should show no effect of ...petition, since it deals with the retrieval of a relatively unencoded stimulus, while other memory ...tems yield results like the "Hebb effect" (Hebb, 1961).

It has recently been suggested however (Dick, 1971) that partial-report techniques are more a source of methodological artifact than evidence supportive of a sensory trace. Dick's findings taken in conjunction with Turvey's (1967) suggest that a more reasonable converging operation would include a full-report condition. If Turvey's results are correct, one would expect a report by repetition interaction, in which a repetition effect occurs in whole report, but not in partial.

Two experiments are reported here in which both full- and partial-report techniques have been used. The second experiment was a replication of the first, with the exception that the nonrepeated slides were composed of letters rather than digits, as in Turvey's (1967) Experiment II.

Method. The *S*s in Experiment I were five male students and one female student solicited from the Department of Psychology at Memorial University. The *S*s in Experiment II were three males and three females from the same population. All *S*s had no previous tachistoscopic experience.

Materials in both experiments were presented by means of a three-field tachistoscope, Scientific Prototype Model GB. The stimulus materials consisted of 96 slides containing three rows of five digits each. These were always dark letters on a white background. No items were repeated more than once within a single array, and never within a single row. In Experiment I the numerals 1–9 were used: in Experiment II the interpolated slides contained letters instead of digits—vowels and the letters Y and Q never appeared.

The exposures were of 100-msec. duration, followed by a 100-msec. delay, at which point a 50-msec.-duration partial-report cue appeared. Pre and post adapting fields were white light, as was the interval between array offset and cue onset. Measurements with a Macbeth illuminometer indicated that the stimulus field had a background value of 5.8 ftL., while the letters were .2 ftL. Pre and post fields were 5.0 ftL. each, while the partial-report cue field had a background value of 5.0 ftL. and the bar marker, .2 ftL. The cue for partial report was a ⅛-in. bar marker which appeared ⅛ in. to the right of the row to be reported. The total array subtended a visual angle of 1.25° horizontally and .75° vertically, while the average letter was .16° high and .14° wide. The intertrial interval was approximately 10 sec.

Each *S* received 96 preexperimental training trials on Day 1. This involved introducing *S*s to the general procedure, establishing facility with the poststimulus cuing procedure, and setting the pace which would be used throughout the experiment. Repeated slides were not used during the training phase. The *S*s were given two kinds of report, full and partial. Under full-report conditions, *S*s were told to write down as many of the items as they could remember, and to place them in the position that they occupied in the display. In the partial-report condition, *S*s were told to write down only those items that appeared in the cued row, again in their correct spatial location. The partial-report condition was a Sperling (1960) arrangement in which cues were quasi-randomly presented, indicating top, bottom, or middle rows of the display. The only constraint was that each row be cued an

[1] This research was supported by university fellowships to the first and third authors, and provincial fellowships to the second and fourth authors. The authors gratefully acknowledge the help of Edward J. Rowe throughout all phases of this work.

[2] Requests for reprints should be sent to Derek Besner, who is now at the Department of Psychology, Reading University, Whiteknights, Reading RG6 2AL, England.

TABLE 1

Mean Percentage Recall

Experiment	Whole report		Partial report	
	Repeated	Nonrepeated	Repeated	Nonrepeated
I	27.80	21.45	32.4	25.4
II	37.4	20.81	42.2	26.75

Note. Partial report scores are calculated on the basis of number correct out of 5; whole report scores are calculated on the basis of number correct out of 15.

equal number of times within a session. Type of report was counterbalanced across *S*s and testing days. Thus half of the *S*s received partial report first on Day 2, and second on Day 3. The remaining half received partial report second on Day 2, and first on Day 3.

On the second and third days, *S*s received 96 trials/day. In Experiment I the repeated slide was presented every second trial; every other trial consisted of a random digit slide. In Experiment II the interpolated slide consisted of letters. Within a daily session, *S*s received 48 presentations of the repeated slide and 48 presentations of the random slides. Under each type of slide, report was half partial, and half full.

Recalled items were scored as correct only when the items were in their exact spatial location. Recall was by means of a written response on a gridded sheet which corresponded to the 3×5 display matrix. Response sheets were turned over before the start of a new trial. Reporting five items correctly under partial report gave a score of 100%; the same number of items under whole report gave a score of 33.33%.

Results. One *S* failed to complete the second experiment—consequently, her data were not included in the analysis. The results were analyzed by a $2 \times 2 \times 6$ within-*S*s analysis of variance in the first experiment and a $2 \times 2 \times 5$ within-*S*s analysis in the second. The respective factors were report (whole or partial), treatment (repetition or nonrepetition), and *S*s. Means for both experiments can be seen in Table 1.

In Experiment I the analysis revealed a main effect of repetition, $F\ (1, 5) = 13.52$, $p < .025$, no effect of report, $F\ (1, 5) = 1.75$, $p > .10$, and no interaction ($F < 1$). The results of Experiment II mirrored those of Experiment I. Repetition was a main effect, $F\ (1, 4) = 20.45$, $p < .025$, there was no effect of type of report $F\ (1, 4) = 1.59$, $p > .10$, and there was no interaction ($F < 1$).

A check on the possibility that the repeated matrix was easier to recall in and by itself yielded negative results. This was done by taking percentage recalled on the first presentation of the repeated slide and comparing it with percentage recalled on the first random slide. In Experiment I the nonrepeated slide was recalled slightly better, although this difference was not significant, Sandler's $A\ (4) = .76$, $p > .10$, two-tailed. In Experiment II the repeated slide was recalled slightly lower than the nonrepeated one, again not significantly, Sandler's $A\ (3) = .31$, $p > .10$, two-tailed, even though the guessing probabilities favor digits over letters by a factor of better than 2:1.

Questioning after Experiment I [illegible]d that all *S*s with the exception of one fail[illegible] notice the presence of a repeated slide. A re[illegible]sis of the data excluding this one *S* left the p[illegible] of results unchanged. In Experiment II all *S*s [illegible]ted noticing that a repeated slide was prese[illegible]

Discussion. These results fail to replicate those of Turvey (1967). However, there are clear differences in methodology; in particular, the contrast ratio between pre and post fields on the one hand and the stimulus field on the other, is much less than in Turvey's. In addition, the partial-report cue was visual rather than auditory, [illegible] stimulus array was much smaller, and the full report condition was an added factor. Finally, only one delay of 100 msec. was used in contrast to Turvey's three. Some of these differences can perhaps be expected to reduce the amount of iconic memory available.

Knowing whether or not the slide was repeated did not appear to systematically influence the manner in which the input was processed, although absolute recall was slightly higher in Experiment II. That is, *S*s in Experiment I who did not know that a slide was repeated appeared to behave in the same fashion as *S*s in Experiment II who did. Furthermore, all *S*s showed the effect in Experiment I, with the exception of one tied score in full report: in Experiment II all *S*s demonstrated the effect under both report conditions.

Since type of report was not a main effect, it is unclear how these results bear on the question of whether iconic memory is susceptible to repetition effects. In general, the existence of excess sensory information is inferred from the superiority of partial over full report, thus it seems likely that these experiments show only that repetition affects a postattentional verbal component rather than an iconic one.

Finally, it may well be that the Sperling (1960) technique is an inappropriate one for this particular question. Reporting items from a visual display entails knowing the names of the items. Retrieval of such information is presumably from a long-term store—thus, in such a paradigm, it would always be possible to argue that repetition effects are dependent on previous excitation of name units in verbal memory rather than upon a precategorical sensory trace.

It is thus suggested that an uncritical acceptance of Turvey's (1967) results, and their implications, would be premature.

REFERENCES

D . . On the problem of selection in short term visual memory. *Canadian Journal of Psychology*, 1971, 25, 3.

H . O. Distinctive features of learning in the higher . . . In J. F. Delafresnaye (Ed.), *Brain mechanisms and* . . . London: Oxford University Press, 1961.

SPERLING, G. The information available in brief visual displays. *Psychological Monographs*, 1960, 74(11, Whole No. 498).

TURVEY, M. T. Repetition and the preperceptual information store. *Journal of Experimental Psychology*, 1967, 74, 289–293.

(Received June 26, 1973)

Journal of Experimental Psychology
1973, Vol. 102, No. 5, 903–905

RELATIONSHIP OF INTERVAL FREQUENCY COUNT TO RATINGS OF MELODIC INTERVALS

THOMAS B. JEFFRIES[1]

Los Angeles City Schools, Los Angeles, California

This study tested *S*s' validity in rating the 24 ascending and descending melodic intervals of the octave, using 116 *S*s' ratings based on their judgments as to how often they had heard or experienced each of these interval types in music. (The ratings had been obtained in an earlier study.) An interval count was taken of the total number of times each of the 24 interval types occurred in the printed scores of 53 current and standard popular vocal pieces. The interval ranks determined from the total count for each interval were correlated with the interval ranks determined from *S*s' mean interval ratings. The results indicated a high degree of validity for *S*s' ratings. Other aspects of the interval count were discussed.

Several studies have provided evidence that *S*s can quite accurately judge the frequency with which words and letters appear in the English language (. . ., Attneave, 1953; Howes, 1954). The purpose of the present study was to determine if *S*s could rate the frequency of occurrence of the different melodic interval types heard in music in much the same way as *S*s rated the frequency of occurrence of words and letters in English as described in the studies cited above. The problem for the study was to determine *S*s' validity in rating the 24 ascending and descending melodic intervals within the octave. The *S*s' ratings were based on their judgments as to how often they had heard or experienced each interval type in music. These rating results were reported in an earlier study (Jeffries, 1972). The criterion measure of validity used in the present study was an interval count taken by the author of the frequency of occurrence of the 24 interval types found in several popular vocal music scores. The melodic intervals ranked according to frequency of occurrence determined from the count were correlated with the intervals ranked on the basis of *S*s' mean interval ratings as to how often they had heard or experienced each of the intervals in music in order to discover with what degree of validity *S*s had rated the intervals.

Another objective of the study was to determine if the ascending interval types occurred in significantly greater numbers in the interval count than the descending interval types.

Jeffries (1972) asked 214 college students comprising two groups to rate the 24 ascending and descending melodic intervals of the octave individually on a 10-point scale. The 98 *S*s in Group 1 rated the intervals presented on tape on the basis of how familiar each interval sounded, while the 116 *S*s in Group 2 rated the same intervals on the basis of how frequently they thought they had heard or experienced each interval type in music. Table 1 presents the familiarity and frequency mean ratings for each interval determined from the rating results.

Results of the study showed that *S*s agreed closely in their familiarity and frequency mean ratings; however, an analysis of variance disclosed a highly significant difference between the ratings, indicating that *S*s had rated the intervals on two distinctly different bases. Results also revealed that *S*s rated the ascending intervals as being significantly more familiar and frequently heard as a group than the descending intervals as a group.

In the present study the ranks for the 24 intervals determined from results of the author's interval count were correlated with the ranks for the same intervals determined from the familiarity and frequency mean ratings given in Table 1.

Method. The interval count consisted of a tally of the number of melodic intervals falling under each interval type found in the printed scores of 53 current and standard popular vocal pieces con-

[1] Requests for reprints should be sent to Thomas B. Jeffries, 10514 National Boulevard, Apartment 323, Los Angeles, California 90034.

TABLE 1

INTERVAL FREQUENCY COUNT, FAMILIARITY-FREQUENCY MEAN RATINGS, AND RECOGNITION ERRORS

Interval	Frequency count	Ratings: Familiarity	Ratings: Frequency	Recognition errors
D M2	1,255	6.05 (13)	5.35 (11)	—
A M2	847	7.10 (5)	6.78 (1)	60 (11)
A m3	409	7.68 (1)	6.56 (4)	136 (6)
D m3	392	7.39 (4)	6.33 (5)	—
D m2	375	7.07 (6)	6.03 (7.5)	—
A m2	368	6.26 (12)	5.44 (10)	64 (10)
A M3	198	6.82 (8)	6.03 (7.5)	116 (7)
D M3	160	6.38 (11)	5.34 (12)	—
A P4	152	7.59 (2)	6.76 (2)	112 (8)
D P4	104	5.60 (16)	4.72 (14)	—
A P5	88	6.90 (7)	6.23 (6)	100 (9)
D P5	73	6.59 (9)	5.97 (9)	—
A M6	45	7.51 (3)	6.69 (3)	191 (4)
A m6	33	5.29 (19)	4.41 (16)	279 (1)
A P8	26	6.58 (10)	5.15 (13)	26 (12)
D P8	26	5.45 (17)	4.57 (15)	—
D m6	26	5.23 (20)	4.21 (17)	—
D M6	21	4.23 (24)	3.82 (22)	—
A m7	17	5.33 (18)	4.07 (19)	213 (3)
A a4	13	5.17 (21)	4.17 (18)	265 (2)
D m7	11	4.93 (22)	4.00 (21)	—
D a4	6	4.91 (23)	3.69 (23)	—
A M7	4	5.65 (15)	3.68 (24)	154 (5)
D M7	1	5.99 (14)	4.05 (20)	—

Note. Abbreviations: A = ascending, D = descending, M = major, m = minor, a = augmented, and P = perfect. Ranks are given in parentheses.

tained in two song collections (*Today's Hits*, undated; *Today's Most Popular Hits*, undated). In selecting popular songs as the source material for the interval count, it was assumed that *Ss* who took part in the interval-rating experiment perhaps were more exposed to this type of music generally than to other music styles.

Ascending and descending intervals were counted from the vocal melodies, not from the piano accompaniment parts. For each song analyzed, the count began with the identification of the first interval of the melody; then successive intervals of the melody were identified and counted one by one. The songs were written in a variety of keys. Melody sections marked with repeat signs were analyzed only once. Intervals considered as enharmonics, for example, the diminished fifth and the double diminished fourth, were tabulated in the count as augmented fourths and minor thirds, respectively, for purposes of limiting the classifications to the 24 interval types. Compound intervals, those greater than an octave, were not included in the count. Repeated tones of the same pitch were tabulated as prime intervals. In all, 6,341 intervals were counted from the scores.

Each of the two song collections contained nearly the same number of songs. To determine internal consistency reliability for the count, the 25 interval types, including the prime interval ranked on the basis of the total frequency count obtained for each interval from one song collection, were correlated with the 25 ranked intervals from the other song collection ($\rho = .96$).

Results and discussion. Table 1 lists the 24 ascending and descending intervals ranked on the basis of the total count obtained [illegible] ach interval. The prime interval count is n[illegible] ded because this interval was not rated in th[illegible] iment.

The frequency interval order o[illegible] le 1 was correlated with the ranked interval [illegible] based on the frequency count ($\rho = .83$, p [illegible] The latter was also correlated with the [illegible] rity interval order of Table 1 ($\rho = .70$, $p <$ [illegible] This difference in correlation was not signif[illegible] $p > .05$).

In the experiment, *Ss* rated t[illegible] nding intervals significantly higher in famili[illegible] nd frequency than the descending intervals. [illegible] uestion now was to determine if the ascending [illegible] als occurred more frequently than the descen[illegible] intervals in the music analyzed. The data of [illegible] rearranged from Table 1 show the compari[illegible] count between the like ascending and de[illegible] intervals. A binomial test was employed to [illegible] differences in count yielding $z = 3.65$, p [illegible] indicating significantly more descending tha[illegible] nding intervals counted in the popular songs.

Ortmann (1937) reported on [illegible] dic interval count taken from 126 songs of the [illegible] mantic composers Schumann, Brahms, and [illegible] rd Strauss. In all, 15,459 intervals were counte[illegible] In Brahms's songs, 44% of the intervals counted [illegible] e ascending, while 56% were descending. In [illegible] he songs of Schumann and Strauss, 48% of th[illegible] tervals were ascending, 52% descending. For th[illegible] ounts of the three composers taken together, th[illegible] descending intervals significantly outnumbered [illegible] he ascending intervals. Here, a binomial test [illegible] $z = 7.69$, $p < .001$.

The interval counts discussed above do not agree generally with *Ss*' ratings in which the ascending intervals were rated significantly higher than the descending intervals. This difference could be attributed to many factors. In considering the interval count findings, perhaps taking more interval counts from a variety of composers' works could provide a more definitive picture regarding frequency of interval occurrence than was provided from these two counts alone. From another standpoint, it is also possible that the author's interval rating

TABLE 2

FREQUENCY COUNT FOR LIKE ASCENDING AND DESCENDING INTERVALS

Interval	Interval count: Ascending	Interval count: Descending
M2	847	1,255
m3	409	392
m2	368	375
M3	198	160
P4	152	104
P5	88	73
M6	45	21
m6	33	26
P8	26	26
m7	17	11
a4	13	6
M7	4	1

Note. Abbreviations: M = major, m = minor, P = perfect, and a = augmented.

procedure lacked some sensitivity in measuring the degree of Ss' familiarity and frequency experience with the intervals. Introducing refinements into the testing procedures might produce rating results more in agreement with the interval count findings given than was found in comparing the interval count results with the mean interval ratings given in Table 1. Both the questions of interval frequency count and the interval-rating procedures need further study.

The data provide some evidence that familiarity of melodic intervals may have some bearing on interval learning. Table 1 shows the total number of errors in interval recognition for each of the 12 ascending melodic intervals made by 24 Ss, these data taken from an earlier study (Jeffries, 1967, p. 189). Table 1 lists the same ascending intervals ordered according to results of the interval count. Correlating the two interval orders produced $\rho = -.53$, $p <$ using a one-tailed test of significance, with these results indicating a significant inverse relationship between interval count and rate of error. The higher was the count for particular intervals, generally, the lower the error rate for Ss learning to identify these intervals. Although the correlation figure above is not particularly high, it does provide some basis for considering further experiments to investigate the precise effects that familiarity may have on interval learning.

In summary, the results showed a close relationship between the interval count findings and Ss' interval frequency ratings. This may point to the possibility that people can judge quite accurately the frequency with which they hear the various melodic intervals in music.

Significantly more descending than ascending intervals were counted in the popular songs as well as in the art songs. These findings are not in agreement with Ss' ratings that the ascending intervals were more familiar and frequently heard than the descending intervals.

Of course, a host of questions was not touched upon in the study, such as those relating to the resolution tendencies of tones, melodic intervals heard in cadences, and the like. No doubt many such factors in one way or another influence the recognition of tones in sequence. While such phenomena may well be of interest to this study, none was considered here because of the limited nature of the problem. Certainly many of these questions call for investigation.

REFERENCES

Attneave, F. Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, 1953, **46**, 81–86.

Howes, D. On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology*, 1954, **48**, 106–112.

Jeffries, T. B. The effects of order of presentation and knowledge of results on the aural recognition of melodic intervals. *Journal of Research in Music Education*, 1967, **15**, 179–190.

Jeffries, T. B. Familiarity-frequency ratings of melodic intervals. *Journal of Research in Music Education*, 1972, **20**, 391–396.

Ortmann, O. Interval frequency as a determinant of melodic style. *Peabody Bulletin*, 1937, December, 3–10.

Today's Hits and Golden Standards. New York: Big 3 Music Corp., undated.

Today's Most Popular Hits. New York: Big 3 Music Corp., undated.

(Received July 2, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 905–908

THE TIME COURSE OF LATERAL ASYMMETRIES IN VISUAL PERCEPTION OF LETTERS[1]

ARNOLD WILKINS[2] AND ANNE STEWART

University of Sussex, Brighton, Sussex, England

A single letter presented at fixation was followed after a short or long interval by a second letter presented in the right or left visual field. The Ss indicated by a manual response whether or not the letter names were the same. For letter pairs with the same name and case, there were fewer errors at short intervals when the second letter in each pair was presented to the left visual field, although at long intervals right visual field presentations were more accurate. For letters with the same name and different case, there were no field differences at short intervals, but at long intervals there was a right visual field superiority. It is proposed that in the process of letter perception a preliminary nonverbal code is available in the right hemisphere until superceded by a verbal code in the left hemisphere.

Cohen (1972) and Geffen, Bradshaw, and Nettleton (1972) found that the matching of letter pairs with the same name and shape is faster when the letters are presented to the left visual field than

[1] This work was conducted while the first author held a Medical Research Council studentship.

[2] Requests for reprints should be sent to A. J. Wilkins, Department of Neurology and Neurosurgery, Montreal Neurological Institute, 3801 University Street, Montreal 112, Quebec, Canada.

TABLE 1

Errors and Latencies for *SAME* and *DIFFERENT* Responses as a Function [illegible] Case Similarity, Interletter Interval (ILI), Visual Field of Second Letter (VF), and Responding Hand

Measure	Response condition	Same case: Short ILI, Left VF	Same case: Short ILI, Right VF	Same case: Long ILI, Left VF	Same case: Long ILI, Right VF	Di[illegible]: Short ILI, Left VF	Di[illegible]: Short ILI, Right VF	Di[illegible]: [illegible] ILI, [illegible]	Di[illegible]: [illegible] ILI, Righ[illegible] VF
Error (%)	*Same*								
	Right hand (Group 1)	0.0	9.4	10.9	0.0	12.5	14.1	[illegible]	0.0
	Left hand (Group 2)	0.0	7.8	7.8	0.0	6.3	7.8	[illegible]	4.7
	Different								
	Right hand (Group 2)	1.6	3.1	3.1	1.6	3.1	9.4	[illegible]	3.1
	Left hand (Group 1)	0.0	0.0	3.1	1.6	4.7	1.6	[illegible]	0.0
Latency of correct responses (in msec.)	*Same*								
	Right hand (Group 1)	732	735	667	707	783	716	[illegible]	743
	Left hand (Group 2)	754	737	686	743	697	700	[illegible]	700
	Different								
	Right hand (Group 2)	649	685	681	657	661	706	[illegible]	670
	Left hand (Group 1)	799	749	730	763	757	768	[illegible]	809

when they are presented to the right. They also showed that matching of pairs with the same name and different shape is more rapid for presentations to the right visual field than to the left. Since each visual half-field projects to the contralateral cerebral hemisphere, the above findings suggest that letters with the same name and shape are matched in the right hemisphere in terms of a spatial code, whereas pairs with the same name but different shape are matched in the left hemisphere on the basis of a verbal representation.

Using a memory drum, Posner and Keele (1967) presented pairs of letters successively and varied the interletter interval. At short intervals pairs with the same name and shape were matched faster than those with the same name and different shape, but at intervals of 1.5 sec. the two types of letter pairs were classified at the same speed. Posner and Keele interpreted their results as indicating that at long intervals all matching was performed in terms of a verbal code, the spatial code having been allowed to decay. If their interpretation is correct, it is conceivable that a visually presented letter is available initially as a spatial code in the right hemisphere until such time as this code is translated into a verbal code in the left hemisphere. The following experiment provides evidence to support this view.

Method. Sixteen right-handed female students at the University of Sussex served as unpaid *Ss*. They were unaware of the purpose of the experiment. The Edinburgh Handedness Inventory (Oldfield, 1971) was administered after the experiment and, on the basis of this questionnaire, all *Ss* were strongly right handed.

The stimuli were presented in a three-field tachistoscope with fields of luminance 34 cd/m², 10.2 cm. high and 15.2 cm. wide, viewed binocularly at a distance of 54 cm. The fixation field, which was illuminated continuously except when one of the stimulus fields was on, contained a [illegible] gray diagonal cross, 2 mm. high, at the cent[illegible] the field. The *Ss* were instructed to fixate the cro[illegible]roughout each trial, and the importance of the[illegible]structions was stressed. The stimulus fields eac[illegible]ontained a single letter approximately 4 mm. high printed in black Letraset 14 pt. Helvetica Light. In the first stimulus field the letter was positioned in the center of the field, and in the second field the letter appeared 28 mm. (approximately 3° of visual angle) to the right or left of center.

A trial began when *E* prepared *S* by counting "Three, two, one," and, on the count of "One," triggered the onset of the first stimulus field. The tachistoscope was wired so that both stimulus fields lasted 100 msec., separated by an interval of either 50 or 990 msec. The *S* was required to respond on one key if the letters had the same name and on a second key if the names were different. Response latencies were measured from the onset of the second stimulus field, and all errors were recorded. Half the *Ss* (Group 1) used the index finger of the right hand for a *same* response and the index finger of the left hand for a *different* response. The remaining *Ss* (Group 2) used the reverse configuration.

Sixteen conditions resulted from the combination of (*a*) the two interletter intervals, (*b*) the similarity or dissimilarity of the letter names, (*c*) the similarity or dissimilarity of the letter cases, and (*d*) the two visual fields in which the second letter of each pair could appear. The experimental conditions are shown in Table 1. The letters B, D, G, H, N, Q, R, and T, or their lowercase equivalents, were used once in each condition as the first letter of a pair so that the case of the initial letter was balanced across conditions. Each cell in Table 1 is therefore based on eight observations from each of eight *Ss*. The second letter in each pair was chosen as appropriate to the condition. For the conditions in which the letter names were different, the second letter

was chosen at random from the same population as the first, with the restriction that pairs such as "Dt," with phonetic similarity, and pairs such as "rn," with similar shape, were avoided. The 128 letter pairs were presented after 10 practice trials, and the order of presentation was random and constant for all *S*s. For half the *S*s from each group the interletter interval was switched so that pairs of letters which were separated by a long interval for half the *S*s were separated by a short interval for the remaining *S*s. As a result, effects of interletter interval cannot be attributed to possible concomitant variation in stimulus discriminability.

Results and discussion. The percentage error in each experimental condition is shown in Table 1. An analysis of variance of the error scores for *same* responses over interletter interval, visual field, case similarity, and responding hand yielded a significant interaction between visual field and interletter interval, F (1, 14) = 34.3, $p < .001$, $MS_e = .311$; a significant effect of case similarity, F (1, 14) = 8.50, $p < .05$, $MS_e = .405$; and a significant three-way interaction between case similarity, interletter interval, and responding hand, F (1, 14) = 6.59, $p < .05$, $MS_e = .267$. The analysis of variance is generally considered a robust test, but in view of the highly skewed distribution of error scores, all the major effects obtained in the analysis were corroborated by individual sign tests with *S*s as *N*.

For letter pairs with the same name and case (e.g., "BB") there were fewer errors at short interletter intervals when the second letter in each pair was presented to the left visual field ($p = .004$, sign test). At long intervals there were fewer errors when the second letter was presented in the right visual field ($p = .004$, sign test). Since the first letter appeared at fixation and was therefore projected to both hemispheres, these results suggest that at short intervals letter pairs which had the same name and case were processed in the right hemisphere on the basis of a spatial representation, whereas at long intervals a verbal match was performed in the left hemisphere.

For letter pairs with the same name and different case (e.g., "bB"), there was no field difference at short interletter intervals ($p = .754$, sign test), although at long intervals fewer errors were made when the second letter was presented in the right visual field ($p = .004$, sign test). The absence of a right visual field advantage at short intervals is of interest because letter pairs with the same name and different case ostensibly required verbal processing at both short and long intervals. A right visual field advantage at both intervals might therefore have been anticipated. Furthermore, inspection of the error scores for these stimuli suggests that the significant three-way interaction obtained in the analysis of variance was partly due to the greater accuracy of the left hand at short intervals and the right hand at long intervals. Both this interaction and the absence of a right visual field advantage when the intervals were short point to the involvement of the right hemisphere. If the retrieval of the name of the first letter was not always complete by the time the second letter was presented and the retrieval of two letter names was not able to proceed simultaneously, a nonverbal code of one letter would have to have been held (possibly in the right hemisphere) while the name of the other letter was retrieved.

An analysis of variance of the error scores for *different* responses over interletter interval, visual field, case similarity, and responding hand yielded one significant effect, that of case similarity, F (1, 14) = 7.47, $p < .05$, $MS_e = .268$. Evidently more errors were made when the pairs differed in case than when their case was the same. It is therefore possible that *same* and *different* judgments involved separate mechanisms (Egeth & Blecker, 1971), and *S*s occasionally responded *different* without retrieving the letter name if the letters were the same size and did not produce a spatial match.

For each *S* the median latency of correct responses in each combination of the two levels of interletter interval, visual field, and case similarity was calculated. The eight medians in each condition were then averaged, and the resulting means appear in Table 1. Analyses of variance based on the median latencies yielded three small effects, two of which tended to contradict the clear pattern which emerged in the analyses of errors. An analysis of the latencies for *same* judgments yielded a significant interaction between case similarity and responding hand, F (1, 14) = 6.05, $p < .05$, $MS_e = 6{,}206$; a similar analysis for *different* judgments revealed an effect of case similarity, F (1, 14) = 7.78, $p < .05$, $MS_e = 2{,}188$, and a three-way interaction between interletter interval, visual field, and responding hand, F (1, 14) = 4.99, $p < .05$, $MS_e = 3{,}310$.

The effect of case similarity for *different* responses indicates that latencies were shorter for pairs of letters with the same case than for letters with different case, and this result is consistent with the view expressed earlier that letter size may have been occasionally used as a shortcut to the production of a *different* response. The interaction between case similarity and responding hand may be attributed to the faster responses with the right hand for letter pairs with the same name and case and the faster responses with the left hand for letter pairs with the same name and different case. This effect and the three-way interaction between interval, visual field, and responding hand are open to many highly speculative interpretations. Instead of discussing these here, it will simply be noted that there was no evidence for a time–error tradeoff since the correlations between residual errors and latencies were positive (.22 for *same* responses and .10 for *different* responses). Error scores were probably a more reliable index of performance than latencies, owing to the use of unpracticed *S*s and the fact that the foreperiod included the variable interletter interval. Cohen (1972) and Geffen et al. (1972), who obtained field differences for latencies but not for errors, used a constant foreperiod, a relatively large number of trials, and, in the

case of Cohen's study, near-threshold stimulus durations.

Although the apparatus did not admit observation of eye movement, presentation of the first letter at the center of the visual field was probably sufficient to secure central fixation at the beginning of each trial. Drift in fixation may have occurred during the long interletter interval and if such drift occurred predominantly to the right, it would provide an alternative account of the general superiority of the right visual field at long intervals. However, the differences in the field advantages which occurred at short intervals cannot be explained in terms of fixation drift. Furthermore, Geffen et al. (1972), who observed eye movements in a situation very similar to that of the present experiment, found that such movements were rare and that they were evenly distributed between the two visual fields. The occurrence of eye movement artifact in the present experiment was therefore unlikely. The development of consistent scanning strategies was also unlikely since *S*s could never predict which visual field the second letter wo[illegible]cupy, which interval would occur, and whethe[illegible]tters would be the same or different shape.

In conclusion, the substantia[illegible]ences in the lateral asymmetries at short an[illegible]ng intervals suggest that in the course of le[illegible] perception a nonverbal code is available in the [illegible] hemisphere until superceded by a verbal c[illegible] in the left hemisphere.

REFERENCES

Cohen, G. Hemispheric differences in a[illegible] classification task. *Perception & Psychophysics*, 1972, [illegible]39–142.

[illegible]h, H., & Blecker, D. Differential [illegible] of familiarity on judgments of sameness and diff[illegible] *Perception & Psychophysics*, 1971, 9, 321–326.

Geffen, G., Bradshaw, J. L., & Nett[illegible] Hemispheric asymmetry: Verbal and spat[illegible] of visual stimuli. *Journal of Experimental Psy*[illegible] 95, 25–31.

Oldfield, R. C. The assessment and ana[illegible] handedness: The Edinburgh Inventory. *Neuropsy*[illegible]*ia*, 1971, 9, 97–113.

Posner, M. I., & Kee[illegible], S. W. Decay of [illegible] information from a single letter. *Science*, 1967, 158, 1[illegible]39.

(Received July 5, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 908–910

EFFECTS OF EVALUATION APPREHENSION ON MEMORY OVER INTERVALS OF VARYING LENGTH[1]

RUSSELL G. GEEN[2]

University of Missouri

In a 2 × 2 × 3 factorial design, 240 female *S*s studied a paired-associate list for a single trial after being told that it was either a test of intelligence or an evaluation of the list. The *S* was either observed or not observed by *E* during the study trial. Recall was tested after 2, 15, or 45 min. Recall was poorer after 2 min. for *S*s who had been told that their intelligence was being evaluated than for *S*s told that the list was being tested, but superior after 45 min. This finding indicates that evaluation apprehension creates a condition of arousal that facilitates long-term memory but hinders memory over relatively short intervals.

In a recently reported experiment, Geen (1973) found that female *S*s who were observed during a single study trial in a paired-associate (PA) learning task subsequently showed poorer recall after 2 min. than nonobserved *S*s, but superior recall after 45 min. Since observed *S*s also reported themselves as more "nervous" than nonobserved *S*s, the results were interpreted as being consistent with the findings of others that arousal during trace formation facilitates long-term retention, while hindering retention over relatively short periods (e.g., Kleinsmith & Kaplan, 1964). The results also conform to those of other investigators who report that the presence of observers produces increased arousal level (e.g., Cottrell, Wack, Sekerak, & Rittle, 1968). Geen implied that being observed causes increased arousal by eliciting an emotional response associated with apprehension over being evaluated by the observer. The present experiment was designed to vary observation and evaluation independently of each other in an orthogonal design. Half of all *S*s were informed that they were being evaluated and half were not. In addition, half of all *S*s were observed by *E* during the PA study trial and half were not. All *S*s were then tested for recall after either 2, 15, or 45 min.

Method. Two hundred and forty female undergraduates served as *S*s in order to earn points applicable to a final grade in a beginning psychology course. The *S*s were selected at random from class lists and scheduled by telephone. Scheduling was

[1] This study was supported by Research Grant GS 2748 from the National Science Foundation.

[2] Requests for reprints should be sent to Russell Geen, Department of Psychology, McAlester Hall, University of Missouri, Columbia, Missouri 65201.

arranged so that one S arrived at the waiting room every 15 min. Before meeting each S, E assigned her to 1 of the 12 conditions by referring to a table of random numbers. The E was a female approximately the same age as S.

After preliminary introductions, E led S to a table in the experimental room on which had been placed a Lafayette memory drum. After explaining that the experiment involved "single-trial learning," E informed S that a PA list consisting of six syllables and associated digits would be shown once and that recall would be tested later. To half of the Ss (evaluation condition), E then said that good recall in PA learning has been shown in several studies to be highly correlated with intelligence and that the experiment was in fact one of a number of pilot studies being carried out to assess intelligence by this means. To the remaining Ss (no-evaluation condition), E said that the experiment was designed as a test of how certain stimuli are recalled better than others and that the session in which S was being run provided for evaluation of one such list.

The E then turned on the memory drum. A period of 10 sec. elapsed before the word READY appeared in the window. Two seconds later the first stimulus term appeared. The stimuli were six syllables of 0% association value (TOV, CEF, QAP, JEX, LAJ, and DAX) first used by Kleinsmith and Kaplan (1964) paired with the digits 2–7. Each syllable appeared alone for 2 sec., followed by the syllable–digit pair for 2 sec., followed by a 2-sec. blank interval before the appearance of the next syllable. Two seconds after the final syllable–digit pair, the word STOP appeared. The S had been instructed to turn off the drum at this point. During the study trial for one half of the Ss (observation condition), E stood just to the left of S and watched as S was presented the syllable–digit pairs. With the other half of the Ss (no-observation condition), E stood beside S at about the same distance as she stood in the observation condition, but kept her back turned to S as she busied herself with some papers on a desk.[3]

When S had turned off the memory drum, E led her to the waiting area outside the experimental room and informed her that she would be called back when it was time for the recall trial. The E began timing the recall interval from the moment S was seated in the waiting room. Each S was then given a short adjective checklist on which she was asked to mark each word that described her state of feeling at that time, on the ostensible grounds that emotions have effects on learning. In addition to several distractor items, the checklist contained the words RELAXED, FEARFUL, JUMPY, and CAREFREE. This was included in an attempt to obtain an independent measure of evaluation apprehension. The E collected the checklist when S had completed it and then informed S how much longer she would have to wait before the recall trial. In order that rehearsal of the items from the PA task might be avoided during the recall interval, each S was then given a questionnaire on vocational interests, presumably as part of a pilot study for another experiment. The Strong Vocational Interest Blank for Women was used. Depending on the length of her wait, S was given either the first part, the first two parts, or the entire blank.[4] When the recall interval was over, E interrupted S, took the questionnaire, and led S back to the experimental room. The E next turned on the memory drum, instructed S to get ready, then watched S and recorded her responses. After a blank interval of 5 sec., the first syllable appeared. Each syllable appeared for 2 sec., followed by a 4-sec. blank interval during which time S responded. The syllables appeared in the same order as in the study trial. After the recall trial had ended, E interviewed each S to discover whether the experimental deceptions had been successful. The S was asked to state whether she attached any significance to whether or not E had watched her during the study trial, whether she believed that intelligence was being tested, and whether the amount of time spent waiting had any effect on recall. Evidence from S's answers that she had formed any hypotheses consistent with those of the experiment was considered grounds for classifying S as "aware" of the deceptions. The E then fully explained the true nature of the study, and included a statement of the major hypotheses being tested.

Results and discussion. No S reported being aware of the true nature of the study.

Significant differences across conditions were found in the number of Ss reporting themselves to be relaxed. No such differences were found in the case of the other adjectives. With an n of 60 in each group, the numbers of Ss in each checking the word *relaxed* on the checklist were: no observation–no evaluation, 42; observation–no evaluation, 28; no observation–evaluation, 10; observation–evaluation, 20. This pattern of responses would not have been expected on the basis of chance, χ^2 (1) = 4.96, $p < .05$, Yates's correction applied. We thus have some evidence that being evaluated made Ss less relaxed than they otherwise would have been.

Figure 1 shows the number of digits recalled by Ss in the 12 conditions of the experiment. An analysis of variance revealed that two first-order interactions were significant, the Evaluation × Time, F (2, 228) = 13.39, $p < .01$, and Observation × Time, F (2, 228) = 3.24, $p < .05$, interactions. The second-order interaction was not significant, F (2, 228) = 2.34, $p < .10$. No other interaction or main effect approached significance. Post hoc t tests further showed significant differences between the evaluated–2-min. and not-evaluated–2-min. groups,

[3] Geen (1973) found that the mere presence or absence of E either during the study trial or the test trial had no effects on S's recall. Accordingly, E was present with S during both trials in the present study.

[4] Most of the time, more than one S occupied the waiting area at one time. In order that the activities of all Ss might be relatively standardized during the recall interval, E instructed each S not to communicate with any other S during the waiting period but to concentrate on filling out the questionnaire. The Ss were also seated in the waiting room in such a way that they did not have face-to-face contact with each other.

t (38) = 2.88, $p < .01$, and the evaluated–45-min. and not-evaluated–45-min. groups, t (38) = 2.26, $p < .05$. The difference between the observed–2-min. and not-observed–2-min. groups failed to attain statistical significance, t (38) = 1.92, $p < .10$. The overall pattern of comparisons among means indicates that *S*s who believed that they were being evaluated manifested superior recall to that shown by nonevaluated *S*s after a relatively long waiting period, but poorer recall following a short period.[6] The mere fact of being observed during the study trial, independent of being evaluated, produced only a weak crossover effect of recall over waiting periods, with practically no difference as a function of being observed or not observed after 45 min. The findings thus corroborate the author's previous report that socially produced arousal promotes memory over a relatively long period but hinders it over a relatively short period, and also indicate that the important variable in the process is apprehension over being evaluated. The delineation of evaluation apprehension as the principal source [illegible] influence on memory is consistent with the fi[illegible] of others regarding the importance of th[illegible] variable (e.g., Paulus & Murdoch, 1971).

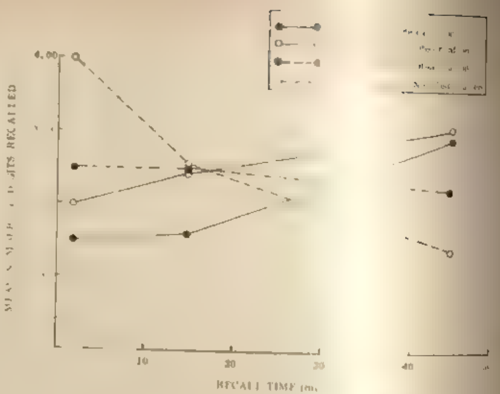


FIGURE 1. Mean number of digits [illegible] treatment groups over different recall [illegible]

[6] Of the 120 *S*s in the evaluation condition, 11 verbalized a lack of belief that they were in fact being evaluated during the postexperimental interview. An internal analysis of the data revealed that these *S*s had recall scores that deviated from those of "unaware" *S*s in the direction expected if we assume that evaluation apprehension produces arousal. Among unaware *S*s, the mean numbers of digits recalled at 2, 15, and 45 min., respectively, were 1.55, 2.19, and 3.33, thus revealing the improved recall over time characteristic of aroused *S*s. Among *S*s aware of the deception, the corresponding means were 3.50, 2.67, and 2.00, exactly what we would expect in the absence of arousal. The small *n*s of the latter three groups (n = 4, 3, and 4, respectively) indicate that the means are not sufficiently stable to warrant statistical analysis. It should also be noted that *S*s aware of the "evaluation" deception were more likely than unaware *S*s to describe themselves as relaxed. 6 of the 11 aware *S*s and 24 of the 109 unaware *S*s did so, χ^2 (1) = 4.03, Yates's correction applied. This finding indicates that *S*s in the evaluation condition were aroused primarily by evaluation apprehension and not by knowledge that they were being deceived.

REFERENCES

COTTRELL, N. B., WACK, D. L., SEKERAK, [illegible] J., & RITTLE, R. H. Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology*, 1968, 9, 245–250.

GEEN, R. G. Effects of being observed on short- and long-term recall. *Journal of Experimental Psychology*, 1973, 100, 395–398.

KLEINSMITH, L., & KAPLAN, S. Interaction of arousal and recall interval in nonsense syllable paired-associate learning. *Journal of Experimental Psychology*, 1964, [illegible] 124–126.

PAULUS, P. B., & MURDOCH, P. Anticipated evaluation and audience presence in the enhancement of dominant responses. *Journal of Experimental Social Psychology*, 1971, 7, 280–291.

(Received July 6, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 5, 910–912

RETRIEVAL BIAS AND THE RESPONSE RELATIVE FREQUENCY EFFECT IN CHOICE REACTION TIME

HAROLD L. HAWKINS,[1] KENNETH SNIPPEL, JOELLE PRESSON, STEPHEN MACKAY, AND DENNIS TODD

University of South Florida

Some investigators have obtained response relative frequency effects in choice reaction time tasks containing many:1 stimulus–response mappings, whereas others have not. Two interpretations of this seeming inconsistency have been proposed. According to one, the effect results from an anticipatory bias in response retrieval which manifests itself only when *S* is permitted the error imbalance which follows as a necessary consequence of the bias. According to the other, the effect results from frequency-related differences in associative strength across response categories whose manifestation is contingent upon the availability of a mediator (e.g., a name) that is strongly associated in *S*'s prior experiences to all stimuli within each response category. The results of the present study are consistent only with the first of these two interpretations.

Hawkins, MacKay, Holley, Friedin, and Cohen (1973) have shown that choice reaction time (RT) is inversely related both to stimulus (S) and response (R) relative frequency when these variables are independently manipulated within a 2:1 S–R mapping task. Both effects were found to increase with declining S–R compatibility. These results were interpreted in terms of an informal statistical

[1] Requests for reprints should be sent to Harold L. Hawkins, Department of Psychology, University of Oregon, Eugene, Oregon 97403. After June 1974, send to Department of Psychology, University of South Florida, Tampa, Florida 33620.

decision model which locates both stimulus and response relative frequency effects within the response retrieval, or selection, process. More specifically, stimulus relative frequency effects were attributed to frequency-related differences in strength of association (and thence, retrieval time) across specific S–R pairings, and response relative frequency effects were attributed to frequency-related bias in response retrieval. However, an inspection of the particular S–R codes used in the Hawkins et al. study raises a serious question regarding the appropriateness of this analysis.[2] Under all compatibility conditions, the two stimuli commonly assigned each vocal-naming response were the upper- and lowercase forms of a single letter. One of the key assumptions underlying the model proposed by Hawkins et al. was that the required response was directly associated with each of the nominal stimuli to which it was assigned. It is equally plausible, however, that this association was mediated by the familiar name common to the two stimuli forming each response category. Were this the case, the response relative frequency effect observed by these investigators may not index retrieval bias but rather frequency-related differences, across response categories, in the strength of association between mediating and required responses. Accordingly, the response relative frequency effect might be expected to attenuate or disappear under conditions where stimuli within categories do not share a common familiar name (cf. Bertelson & Tisseyre, 1966; Hawkins & Friedin, 1972; Hawkins, Thomas, & Drury, 1970). The present experiment constituted a test of this implication.

Method. The *S*s were 14 men and 10 women enrolled in introductory psychology at the University of South Florida. All *S*s had normal or normal corrected vision, and all were naive with respect to RT research. Extra course credit was given for participation.

Stimuli consisted of the letters D, d, Q, q, H, h, k, b, and p (Tactype No. 5418). The manner of stimulus presentation was as reported in Hawkins et al. (1973).

All *S*s were pretested on a 40-trial practice task in which a colored nonsense syllable, either green (p = .50) or orange, was presented on each trial. The vocal responses during practice were the common names of the two colors. Based on performance on this task, *S*s were blocked into slow (N = 12) and fast responders. One-third of the *S*s in each block were assigned on a random basis to one of three levels of S–R compatibility. The task, under all levels of compatibility, was to vocalize one of three responses to each of six perceptually distinct stimuli. Stimuli and responses were paired in 2:1 fashion. The two stimuli within one response category were presented with relative frequencies .50 and .10, respectively; those within another response category both were assigned a relative frequency of .10; and the remaining two stimuli were presented with a relative frequency of .10. These latter two stimuli, together with their respective response, were distractors, included to increase overall task difficulty and to minimize the likelihood *S*s would employ a "positive set" strategy of the type observed by Hawkins and Hosking (1969) in two-response choice RT tasks. Of primary interest were the high- (HH) and low-frequency (LH) stimuli requiring a common high-frequency (.60) response and the two low-frequency (LL) stimuli requiring a common low-frequency (.20) response.

All *S*s were tested on a total of 240 trials during a single experimental session lasting about 1 hr. and occurring within 3 wk. of the practice session. During the experimental session, the identity of, the correct relative frequency of, and the response appropriate for each stimulus were made constantly available to *S* on an index card. The intertrial interval was 3.5 sec., with a warning tone of 1-sec. duration occurring immediately prior to stimulus onset. After each correct response, *S* was informed of his RT in milliseconds. Following each incorrect response, *S* was told that an error had been committed. Stimuli to which *S* incorrectly responded were not subsequently reported. Short breaks were given following each 40-trial block. Under all experimental conditions, *S*s were instructed to respond as rapidly as possible while at the same time, minimizing errors.

Eight *S*s were tested under the high-compatibility condition where they were required to vocalize the name commonly associated with each of the stimulus letters. Each of the four stimuli, D, d, Q, and q, served as the high-frequency alternative for two of the eight *S*s. The two stimuli H and h served as distractors. Eight *S*s were tested under a low-compatibility condition where they were required to respond *Z* to stimuli D and d, *V* to Q and q, and *J* to distractors H and h. D, d, Q, and q each served as the high-frequency alternative for two of the eight *S*s. The remaining eight *S*s were tested under a low-compatibility condition where the response *Z* was required to either D or k, the response *V* was required to either Q or b, and the distractor response *J* was required to either H or p. As before, D, k, Q, and b each served as the high-frequency alternative for two of the eight *S*s.

Results and discussion. Table 1 gives the mean correct RT and percent errors (in parentheses) under each of the conditions of the experiment. These data were subjected to a mixed model analysis of variance with compatibility (high vs. low$_1$ vs. low$_2$) and blocks (fast vs. slow *S*s) treated as between-*S*s variables and S–R pairing (HH vs. LH vs. LL) and level of practice (first through fourth block of 60 trials) treated as within-*S*s variables. The overall analysis revealed that S–R compatibility, F (2, 18) = 101.99, S–R pairing, F (2, 36) = 116.51, the Compatibility × S–R pairings interaction, F (4, 36) = 34.18, practice, F (3, 54) = 29.20, the Practice × Compatibility interaction, F (6, 54) = 8.06, the S–R Pairing × Practice interaction, F (6, 108) = 7.78, and the second-order S–R Pairing × Compatibility × Practice interaction, F (12, 108) = 5.01, were all significant sources of variance (p < .001).

It will be recalled that the primary objective of the present study was to determine whether the

[2] The authors wish to thank Steven W. Keele and Lester E. Krueger who independently alerted us to this possibility.

TABLE 1

Mean Choice Reaction Time and Percent Errors (in parentheses) Ac[illegible] Conditions and Levels of Practice

Trial block	High HH	High LH	High LL	Low (common name) HH	Low (common name) LH	Low (common name) LL	Low (di[illegible] names) HH	Low (di[illegible] names) LH	Low (di[illegible] names) LL
I	601	648	652	754	866	1013	819	[illegible]	1670
	(.5)	(1.8)	(2.0)	(1.1)	(8.3)	(20.7)	(1.9)	[illegible]	(30.1)
II	573	604	602	678	775	916	786	[illegible]	1276
	(.4)	(1.5)	(1.8)	(1.0)	(6.8)	(16.9)	(2.1)	[illegible]	(25.0)
III	555	593	605	655	734	870	719	[illegible]	1164
	(.4)	(1.4)	(2.0)	(.7)	(5.4)	(15.0)	(1.6)	[illegible]	(23.3)
IV	563	581	598	668	706	821	696	[illegible]	978
	(.6)	(1.5)	(1.8)	(.9)	(6.0)	(16.3)	(1.1)	[illegible]	(19.8)
Mean across blocks	573	607	614	689	771	905	755	[illegible]	1272
	(.5)	(1.6)	(1.9)	(.9)	(6.6)	(17.2)	(1.7)	[illegible]	(24.6)

Note. HH and LH refer to high- and low-frequency stimuli, respectively, requiring a common high-frequency [illegible] response, LL refers to the two low-frequency stimuli requiring a common low-frequency (.20) response.

response relative frequency effect attenuates or disappears when stimuli within response categories do not share a common name in *S*'s prior experience. Two analyses were performed which are relevant to this question. First, Fisher's least significant difference (LSD) test, with the criterion of significance set at .05, was used to assess the occurrence of response relative frequency effects under each of the three compatibility levels. This analysis revealed that choice RT increased significantly between LH and LL S-R pairings under both low-compatibility, but not under the high-compatibility, conditions (LSD = 34 msec.). Second, a comparison of the magnitudes of the response relative frequency effect across the two low S-R compatibility conditions revealed that the effect was significantly *greater* under the condition in which stimuli comprising each response category did not share a common familiar name, $F(1, 36) = 7.85$, $p < .01$. These results would seem to be consistent with the Hawkins et al. (1973) view that the response relative frequency effect indexes anticipatory bias in response retrieval and that such anticipatory adjustments represent a feature of the information-processing sequence which extends across a variety of speeded recognition tasks. However, a further feature of the present data indicates that such an account cannot, in and of itself, fully characterize the findings reported here. Specifically, a subsequent partition of the variance in the original analysis revealed that the magnitude of the Response Relative Frequency × Compatibility interaction declined with practice, $F(6, 108) = 4.03$, $p <$ .005. This interaction reflects the [illegible]hat the RT difference between LH and LL pairings declined with practice under the low-compatibility condition where stimuli within response cat[illegible]s did not share a common name, but not under [illegible] remaining two compatibility conditions. The anti[illegible]patory bias interpretation implies, to the contrary, that since *S*s were given equal practice on LH and LL pairings under all compatibility conditions, the practice-related declines in RT for these two S-R pairing types should have remained equal across trial blocks, yielding no decline in the response relative frequency effect under any of the compatibility conditions.

The present results make it clear that the response relative frequency effect is a complicated phenomenon which may prove to be difficult or impossible to adequately characterize in terms of a single factor model of the type proposed by Hawkins et al. (1973).

REFERENCES

Bertelson, P., & Tisseyre, F. Choice reaction time as a function of stimulus versus response relative frequency of occurrence. *Nature*, 1966, **212**, 1069–1070.

Hawkins, H. L., & Friedin, B. D. The relative frequency effect and S-R compatibility. *Psychonomic Science*, 1972, **28**, 329–330.

Hawkins, H. L., & Hosking, K. Stimulus probability as a determinant of discrete choice reaction time. *Journal of Experimental Psychology*, 1969, **82**, 435–440.

Hawkins, H. L., MacKay, S. L., Holley, S. L., Friedin, B. D., & Cohen, S. L. Locus of the relative frequency effect in choice reaction time. *Journal of Experimental Psychology*, 1973, **101**, 90–99.

Hawkins, H. L., Thomas, G. B., & Drury, K. B. Perceptual versus response bias in discrete choice reaction time. *Journal of Experimental Psychology*, 1970, **84**, 514–517.

(Received July 12, 1973)

Journal of Experimental Psychology
1974, Vol. [illegible], No. 5, 913–914

OVERT REHEARSAL AND LONG-TERM RETENTION

GARY F. MEUNIER,[1] JANE KESTNER,[2] JO ANN MEUNIER, AND DOUGLAS RITZ[3]

Ball State University

Recall scores from intentional and incidental learning groups were compared when rehearsal was made overt by having *S* write each stimulus as it appeared. Duration of writing time was varied and the number of written replications of a stimulus was an index of rehearsal. Intentional learning *S*s were clearly superior to incidental learners even though the groups rehearsed equally often. In addition, moderate increases in rehearsal time led to improved retention, but increases in rehearsal beyond this level did not affect recall.

A central aspect of most multistore models of memory is the notion that stimulus material is both maintained in a limited capacity, temporary storage and transferred to a larger, more durable storage by a process of rehearsal. And, generally, the greater the amount of rehearsal given an item the greater is the probability that the item will be encoded into the more durable store (e.g., Atkinson & Shiffrin, 1968).

A testable implication of this conceptualization is that, all other factors constant, long-term retention of a given item should be a function of the amount of rehearsal it receives. Unfortunately, attempts to evaluate this statement have not provided consistent results: Using a variety of paradigms, previous research has yielded evidence both supportive (e.g., Rundus & Atkinson, 1970) and contradictory (Stanners & Meunier, 1969) to the implication. The present experiment, thus, is an attempt to further clarify the relationship between rehearsal and retention.

A "rehearsal" in this experiment consists of a written replication of the stimulus item. Although there is, of course, some controversy over just what a rehearsal is and how it is to be measured, this definition does not seem inconsistent with other frequently used conceptions and does have the advantage of making the rehearsal process overt.

Method. The *S*s were 130 students from introductory psychology classes at Ball State University, who received extra course credit for participation in the experiment.

Stimulus items were 24 29% association value trigrams selected from the Witmer (1935) scaling. Two additional trigrams selected from the same group served as practice items. The stimuli were projected individually by a Kodak Carousel projector onto a screen located approximately 2 m. in front of the seated *S*. Height of the individual letters on the screen was 7.62 cm., and each was presented for 1.5 sec. The *S*s were run in groups of from three to five members each. Each *S* silently viewed the trigram on the screen while it was presented. Termination of this visual stimulus was the signal for *S* to start writing the trigram on a sheet of paper from a booklet provided by *E*. Writing continued until a verbal command of "Stop" was given by *E*. This command was also the signal for *S* to turn to the next sheet of paper and get ready for the next trigram which was presented in approximately 2 sec. Following presentation of all 24 stimulus items, *S* was asked to recall as many of the trigrams as he could. This recall was written and terminated after 5 min.

The *S*s were divided into two groups dependent upon instructions given before the experiment. All were told that the intent of the experiment was to measure speed and accuracy of writing and how these changed with fatigue. In addition, *S*s in the aware (A) condition were told that they would be asked to recall the trigrams after the experiment was over. No attempt was made to emphasize the importance of this recall over the writing task. Those in the unaware (UA) condition were not told of the final recall until they had been presented each of the trigrams in the writing part of the experiment. Also, the duration of the writing task following each trigram was either 3, 9, or 15 sec. and was varied randomly with the constraint that each interval was presented equally often (eight times) within a session. All members of a given session were given either A or UA instruction, and a total of 65 *S*s served in each of the two groups.

Results. The results may best be considered in two parts, an analysis of the number of rehearsals of a trigram, and an analysis of the trigram recalls.

One rehearsal resulted when all three letters of a trigram were written in the correct serial order. These results are presented in Table 1 where both means and standard errors are given for each of the

TABLE 1

MEANS AND STANDARD ERRORS (IN PARENTHESES) OF NUMBER OF REHEARSALS PER ITEM AS A FUNCTION OF REHEARSAL INTERVAL (IN SEC.) AND S'S AWARENESS OF FINAL RECALL

Group	Rehearsal interval 3	9	15
Aware	2.42 (.05)	5.66 (.10)	8.42 (.16)
Unaware	2.48 (.07)	5.76 (.12)	8.89 (.16)

[1] Requests for reprints should be sent to Gary F. Meunier, Department of General and Experimental Psychology, Ball State University, Muncie, Indiana 47306.
[2] Now at the University of Notre Dame.
[3] Now at the University of Arkansas.

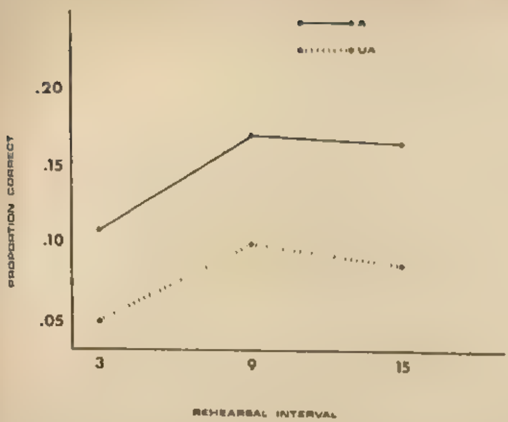


FIGURE 1. Proportion of items recalled as a function of rehearsal interval and S's awareness of final recall. (Abbreviations: A = aware and UA = unaware.)

six treatment combinations. An analysis of variance of these data confirmed the obvious conclusions from the table. The only significant effect was attributable to length of the writing interval with an increasing number of rehearsals given each item as the length of the interval increased, F (2, 236) = 115.54, $p < .001$. The *S*s in the A and UA groups did not differ in the number of rehearsals given nor in the number of rehearsals as a function of interval length. The similarity of the two groups is rather striking and implies that *S*s in the A group did not deemphasize the importance of the writing task.

In order to maintain consistency with the scoring of rehearsals, a correct recall was scored when all three letters of the trigram were recalled in the correct serial positions. The *S*s were, however, instructed to recall the whole trigrams in any order they chose, and a graphical display of these results is presented as Figure 1. An analysis of variance confirmed the suggestions from this figure: Recalls were significantly higher for A than for UA *S*s, F (1, 128) = 33.212, $p. < .001$, and recall probability changed with increasing rehearsal time, F (2, 256) = 8.530, $p. < .001$. Thus, although recall was clearly superior for A *S*s at each rehearsal interval, the same pattern of effects of the rehearsal intervals was found in both A and UA recalls. Increasing rehearsal time from 3 to 9 sec. increased recall substantially, but additional increases in this time beyond 9 sec. did not materially effect recall.

Discussion. Three aspects of these results seem particularly noteworthy. First, *S*s rehearsed equally often regardless of instructions concerning the recall task. Although it is possible tl[...] *S*s were also rehearsing covertly (and thereb[...]reasing their actual rehearsal time) this possib[...]oes not seem likely. The very short interv[...]een *E*'s command to stop writing and th[...]rance of the next trigram would not allow m[...]e for covert rehearsal, and the equality of [...]hearsals for A and UA *S*s implies that A [...] not doing something in addition to writin[...]rge number of studies indicates that perfo[...] diminishes when two activities are attemp[...] once (e.g., Monty, Taub, & Laughery, 1965

The second salient feature of [...] results concerns the relationship between an[...] of rehearsal and retention. While the numbe[r] of rehearsals increased by significant amounts [...]ch increase in rehearsal time, a correspondin[g in]crease in recall was found only when writing time [in]creased from 3 to 9 sec. The extra rehearsals allo[wed] by lengthening this time to 15 sec. did not prod[uce] increases in recall. While this breakdown in the positive relationship between memory and rehearsal appears to contradict a great deal of experimental literature (e.g., Rundus & Atkinson, 1970), it is not totally unprecedented. For example, Stanners and Meunier (1969) have also reported a limit to the beneficial effects of rehearsal on memory. Most of the studies that have reported a strong positive relationship between memory and rehearsal of individual items have not studied rehearsal intervals greater than 10 sec. or retention intervals longer than 30 sec.

And, third, some process other than simple rehearsal must be invoked to account for recall differences between A and UA *S*s. Although these two groups were nearly identical in terms of rehearsals given, A *S*s yielded consistently superior recall scores. Thus, while the present experiment does not imply what this extra process may be, it does seem to clearly indicate its existence.

REFERENCES

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 2. New York: Academic Press, 1968.

MONTY, R. A., TAUB, H. A., & LAUGHERY, K. B. Keeping track of sequential events: Effects of rate, categories, and trial length. *Journal of Experimental Psychology*, 1965, 69, 224-229.

RUNDUS, D. J., & ATKINSON, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 99-105.

STANNERS, R. F., & MEUNIER, G. F. Pronunciability and rehearsal time in short-term memory with controlled acquisition. *Journal of Experimental Psychology*, 1969, 80, 359-363.

WITMER, L. R. The association-value of three-place consonant syllables. *Journal of Genetic Psychology*, 1935, 47, 337-360.

(Received July 16, 1973)

Journal of Exp[illegible] Psychology
1974, Vol. 1[illegible] 5, 915–916

EXTENDED MEDIATION IN CHILDREN'S PAIRED-ASSOCIATE LEARNING[1]

JUDITH E. SIMS-KNIGHT[2] AND LEWIS P. LIPSITT[3]
Brown University

In a study of paired-associate learning in 80 third graders, both the mediated facilitation and interference effects previously demonstrated on List III by Wismer and Lipsitt were replicated under conditions in which the stimulus words of List III were replaced either with pictures of the stimulus objects or pictures of objects in the same class as the stimulus objects. The results indicated that third-grade children apparently have the ability to make implicit associations in a mediational chain.

[illegible]mer and Lipsitt (1964) demonstrated that fi[illegible]-grade children exhibit both facilitation and interference in a mediational paired-associate study. The *S*s were presented printed word pairs on three [illegible]essive lists to arrange facilitation and interference conditions on a within-*S*s basis: (*a*) A–B, B–C, A–C for facilitation, and (*b*) A–B, B–C, A–C (re-paired) for interference. One control (C_1) involved a new word in List II, so that the sequence was A–B, X–C, A–C, and another control (C_2) instituted new word pairs on List III. Besides demonstrating the predicted facilitation and interference effects, error analyses revealed that the mechanism underlying both types of transfer was mediational in nature.

Russell and Storms (1955) showed that adults can provide their own mediational links when these are not directly administered. The present study attempted to determine whether young children can make implicit chains involving images and words. Two levels of extended mediation were studied. In the first level, pictorial presentations (designated A′) were introduced in List III in place of the printed [illegible] words, e.g., instead of the word *telephone* appearing, a drawing of a telephone would appear. For *S*s to respond to these pictures *as if* they were the printed words of List I presupposes an implicit mediational link between the image and the word.

In the second condition, the List III picture (now called A″) was of a different object from A, but was in the same class as A. For example, if the stimulus word A was telephone, a picture of a microphone (A″) was used in List III. This condition extends the mediational chain still further.

Method. The stimulus words were typed on 4 × 6 in. white cards, and the stimulus pictures were black-and-white line drawings on similar cards. Stimulus cards were presented manually, at a rate of 5 sec. per card, followed by cards containing both the stimulus and response words. On List III pairs, line drawings replaced the printed stimulus words.

Each *S* was presented eight word-pair sequences, two sequences for each experimental condition—facilitation, interference, C_1, and C_2. The eight word-pair sequences were chosen for their high A–B and B–C associative strengths and low A–C associative strength. Word-pair sequences were completely counterbalanced across *S*s. For example, a bull–cow (A–B), cow–milk (B–C) chain appears as a List III facilitation pair in Group 1, as interference in Group 2 where bull is paired with an extramediational word, and as a different type of control in Groups 3 and 4, respectively (see Wismer & Lipsitt, 1964).

The *S*s were 80 third-grade children, 40 in each implicit chain group (A′ or A″). Each counterbalancing group had 10 *S*s, who were tested individually in one session. Standard paired-associate learning instructions preceded 3 trials of List I, 3 of List II, followed by repetitions of List III until *S*s reached a criterion of 3 errorless trials (or 12 trials were completed). Presentation order was randomized separately for each trial.

Results. Table 1 shows that mean errors on List III varied with pair type. This effect was significant in a three-way analysis of variance with pair type as a within-*S*s variable, F (3, 216) = 34.96, $p < .01$. As predicted, the mediated facilitative and interfering effects were greater in the A′ condition than in the A″ condition, yielding a significant interaction, F (3, 216) = 5.44, $p < .01$. Simple tests revealed that in Condition A′ the facilitation pairs produced significantly fewer errors than C_1, t (39) = 2.82, $p < .01$, but not fewer than C_2, t (39) 1.70. Performance under the A′ condition on interference pairs was significantly poorer than performance on both controls, t (39) = 3.42 and 5.60, re-

TABLE 1

MEAN ERRORS ON LIST III FOR FACILITATION, INTERFERENCE, AND CONTROL CONDITIONS UNDER TWO LEVELS OF EXTENDED MEDIATION

Condition	Extended mediation levels: A′	A″
Facilitation	2.00	2.43
Control 2	2.30	2.45
Control 1	2.88	3.00
Interference	4.58	3.53

[1] This study was conducted at Brown University as a senior honor's thesis (Sims, 1965) by the first author under the direction of the second. The authors wish to express their gratitude to Florence McGwin of the Summit Avenue School, Providence, Rhode Island, and Edward Staltare of the Pleasant Street School, Seekonk, Massachusetts, for their kind cooperation.

[2] Now at Wheaton College.

[3] Requests for reprints should be sent to Lewis P. Lipsitt, Walter S. Hunter Laboratory of Psychology, Brown University, Providence, Rhode Island 02912.

TABLE 2

PERCENT FIRST-TRIAL RESPONSES ON LIST III TO A′ AND A″ STIMULI

Condition A′					
B	C	A″	C re-paired (Correct response)	A (labeled picture)	Other response (including none)
41.2	9.9	0	0	10.0	38.9
Condition A″					
B	C	A′	C re-paired (Correct response)	A (labeled picture)	Other response (including none)
6.4	1.8	3.5	0	20.2	68.1

spectively, $p < .01$. In Condition A″, facilitation pairs performance was again significantly better than performance on C_1, t (39) = 2.22, $p < .05$, but not better than performance on C_2, t (39) = .11. Interference-pairs performance was significantly poorer than C_2, t (39) = 2.66, $p < .01$, but not poorer than C_1, t (39) = 1.30.

A significant Pair Type × Counterbalancing Group interaction, F (9, 216) = 2.36, $p < .05$, was due apparently to the relatively poorer performance on Counterbalancing Group 4, particularly on C_1 and interference pairs.

Analysis of first-trial errors (Table 2) indicated that in the A′ condition the most frequent response to A′ was B, while in the A″ condition the most frequent first-trial responses were either to name the stimulus or refrain from responding. Moreover, the A′ condition elicited more than five times as many C (mediated) responses than A″, an effect which undoubtedly contributed to the greater facilitative and interference effects under Condition A′.

Discussion. Third-grade child[illegible]arently have the ability to make implicit asso[illegible] in a mediational chain. All results from th[illegible]ndition indicate that they can make associati[illegible]tween a word and its pictorial representation a[illegible] that association to complete a mediational c[illegible] Second, *Ss'* most frequent response of B to th[illegible]resentation of A′ suggests that they had asso[illegible] A′ with A, which had previously been paired [illegible] B.

The results of the A″ conditio[illegible]pected and as revealed by the Chain Cond[illegible] Pair Type interaction, indicated that as the [illegible]ional chain is lengthened, both facilitative [illegible]nterference transfer are attenuated. Further[illegible] *Ss'* first response to A″ was most often a la[illegible]he picture rather than B or C.

The difficulty of obtaining a p[illegible] control is evident. Wismer and Lipsitt's (19[illegible]ndings had suggested that C_3 more clearly d[illegible]tiates between facilitation and interferenc[illegible]he present study, however, neither condition [illegible] a significant difference between facilitation [illegible], suggesting that even in the absence of exp[illegible] A-B, B-C training *S*s may have been able to [illegible]de the implicit B link, thus facilitating A–C lea[illegible]ing. Since C_1 was significantly different from bo[illegible]acilitation and interference in Condition A′, it [illegible]ay be considered, at least in this study, the b[illegible]er control. Considering only this control, facilit[illegible]on but not interference was obtained under Condit[illegible] A″.

REFERENCES

R[illegible], W. A., & STORMS, L. H. Implicit verb[illegible] in paired-associate learning. *Journal of Experimental Psychology*, 1955, 49, 287–293.

SIMS, J. Extended mediation in children's paired-associate learning. Unpublished honor's thesis, Brown University, 1965.

WISMER, B., & LIPSITT, L. P. Verbal mediation in paired-associate learning. *Journal of Experimental Psychology*, 1964, 68, 441–448.

(Received August 2, 1973)

JOURNAL OF EXPERIMENTAL PSYCHOLOGY
MONOGRAPH

Vol. 102, No. 5, 917–939 MAY 1974

THE ROLE OF THE ASSOCIATION IN RECOGNITION MEMORY[1]

BENTON J. UNDERWOOD[2]

Northwestern University

The purpose of the eight experiments was to assess the role which associations between two words played in recognition decisions. The evidence on weak associations established in the laboratory indicated that the association was playing a small role, but that recognition performance on pairs of words was highly predictable from frequency information. However, the use of strongly associated words indicated that the strength of the association per se was not a critical variable in recognition performance. A post hoc expansion of frequency theory was proposed. Some unexpected findings included criterion differences in making frequency judgments as compared with recognition decisions, and criterion differences in recognition tests on homonym pairs as compared with other classes of word pairs.

At the empirical level, the present studies were concerned with the role that an association between two words plays in the recognition performance for the pair. At the theoretical level, the studies were prompted by a theory which states that recognition decisions are mediated by frequency information (Underwood, 1971). This theory has evolved from work in which the unit of analysis was the single word; the association between words has received little attention in the development of the theory. For the single word, the theory, as applied to recognition decisions between old and new words, simply assumes that these decisions are made by discriminations of phenomenal frequency differences between the old and new words. Associations enter this formulation in only a vague way. It is known that the background frequency of words (as indexed by word counts) has little if any effect on judgments of situational frequency, i.e., frequency induced in the laboratory (Underwood, Zimmerman, & Freund, 1971). This must mean that the frequency information induced in the laboratory is kept distinct from background frequency and, therefore, in the vague sense noted above, situational frequency is somehow associated with the laboratory context.

The broad perspective of memory into which frequency theory fits assumes that various types of information may constitute a memory for an event (Underwood, 1969). Frequency information is simply one of these types, but a type that assumes (by theory) the major function in recognition decisions involving single verbal units. The theory has never presumed that other types of information will always be irrelevant for recognition. Indeed, it has been concluded that when frequency information becomes invalid for recognition decisions the subject will turn to other types of information in memory to mediate his judgments (Underwood & Freund, 1970). A recent study by Broder (1973) shows a serious breakdown of the theory when frequency information and associative information are in conflict. So, then, frequency theory can only be viewed within a larger perspective; the theory deals with one type of information in memory and

[1] This research was supported by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract N00014-67-A-0356-0010, Contract Authority Identification NR-154-321.

[2] Requests for reprints should be sent to Benton J. Underwood, Department of Psychology, Northwestern University, Evanston, Illinois 60201.

its role in recognition performance. It is a part of the theoretical–empirical task to establish the limits of the theory, limits at which the theory is no longer useful in predicting the outcome of recognition studies. The studies to be reported were undertaken to determine if associative information between words leads to a breakdown of the predictions based only on frequency information.

There is a rapidly growing literature dealing with word recognition and the associative relationships among the words. These studies are often spoken of as studies dealing with associative context. The outcomes of these experiments are presumed to be relevant to the question of whether or not recognition involves retrieval. The idea is that if an association influences recognition, retrieval processes *are* involved. The central method of study, with several variations, is that of removing, on the test, the items which occurred with the target word during study. If performance falls (as compared with the condition in which the context remains on the test) it is said to implicate the loss of retrieval cues. Context may be added also, but if this influences recognition, the theoretical interpretation is not clear with regard to the role of retrieval in recognition.

An examination of the literature shows no consistency in the effect of associative context on recognition. This confusion may be sampled. Tulving and Thomson (1971) reported that both the addition of context and the removal of context degraded recognition. Subsequently, Thomson (1972) found little effect of context addition. The data of Cofer, Segal, Stein, and Walker (1969) showed that removing context caused a fall in recognition scores. Light and Carter-Sobell (1970) found that context change did influence recognition (performance fell) but in a further study Light (1972) reported that recognition for words embedded in sentences for study but tested alone gave recognition scores that were quite comparable to those obtained when the word was studied alone and tested alone. Ellis and Shumate (1973) showed no loss in recognition of stimulus terms from a paired-associate list over 28 days although the ability to recall the response terms fell over this period. Wood (196[illegible]) also concluded that the association developed in paired-associate learning did not influence recognition, although Wolford (1971) interpreted his results to indicate that the association between words in a pair was involved in recognition.

It is not the purpose here to attempt to evaluate the possible reasons for the contradictory findings. The details of some of these studies will be considered in conjunction with the specific experiments to be reported. It is merely noted here that the empirical facts concerning the role of associations in recognition decisions are far from clear.

One further background comment is required. It was noted earlier that the studies to be reported were undertaken with frequency theory as the background orientation. This theory fits nicely into the conceptual notions of signal-detection analysis. If, for example, an old and new word are given on a forced-choice test of recognition, the apparent frequency of the old item constitutes the signal, the apparent frequency of the new word, the noise. Words which have the same input frequency will constitute a distribution of phenomenal frequency for old words, and the phenomenal frequency of new words will constitute a distribution of noise. Two presumably independent measures of behavior may be obtained from signal-detection analysis. One of these is discriminability or sensitivity between the signal and noise distribution (between the phenomenal-frequency distributions for old and new words in terms of frequency theory). The measure d' is used to express sensitivity differences. The other measure has to do with the criterion set by the subject for accepting old words and rejecting new ones. Presumably the criterion represents a form of response bias which may differ among subjects and among conditions of an experiment, and it is measured by *beta* in signal-detection analysis. As will be seen, some unexpected differences in the criterion settings or response biases

were found. However, while accepting the conceptual notions of signal detection, it has seemed unnecessary to use the measures of signal detection (d' and *beta*) to reflect the behavioral manifestations of these concepts. Rather, other measures have been developed. Nevertheless, the relationships between the measures used and d' and *beta* will sometimes be reported.

The eight experiments to be described divide themselves into two subsets. The first five were concerned with transient or weak associations developed in the laboratory between the two words in a pair. The remaining three used pairs of words already strongly associated by cultural usage.

Experiment I

The first experiment was planned as an analogue of a previous study using single words (Underwood, 1972, Experiment 1). In that study, comparative recognition decisions were evaluated in conjunction with comparative judgments of frequency under exactly the same conditions. The outcomes were statistically equivalent. Furthermore, when errors were made they occurred for the same items for the group making recognition decisions and for the group making frequency judgments. The correlation between the error distributions was .74. This finding was deemed consistent with the theory that recognition decisions are based on frequency information. In Experiment I, reported below, *pairs* of unrelated words were used as the recognition unit, with frequency judgments and recognition decisions being made by different groups. The empirical question is whether or not the recognition decisions are predictable from judgments of frequency as was true when single words were the recognition unit.

Method

The general method may be described first. The *Ss* were presented a series of pairs for study, each pair being presented on a slide for 5 sec. Before presenting the pairs the *Ss* were given a booklet, and the cover sheet of the booklet contained the test instructions. These instructions were studied by the *S* before the list was presented so that he knew exactly how he would be tested. One set of instructions informed the *S* that he would be asked to make recognition decisions, both absolute recognition (*yes–no*) decisions for individual pairs, and comparative or forced-choice decisions between two pairs, one old and one new. The two types of tests were illustrated on the instruction sheet. Another set of instructions, given to other *Ss*, informed them that their memories would be tested for frequency information. These instructions further indicated that there would be two types of test items. In one case, single pairs were shown and the *S* was required to fill in a blank with a number to indicate the number of times he thought the pair had been presented. In the other case, two pairs were shown and the *S* was required to encircle the pair which had been presented most frequently. Thus, all *Ss* received the same study list, and all *Ss* made both absolute and comparative judgments, but for one group the instructions emphasized correct recognition and for the other group, correct frequency information.

Materials. The pool of words consisted of 240 five-letter words with Thorndike–Lorge frequencies falling between 11–30. This pool consisted of approximately 90% of all five-letter words falling in this frequency range.[4] From this pool, 72 pairs were formed randomly subject only to the restriction that the two words in a pair did not have the same initial letter. The 72 pairs were used in the study list, with 48 pairs presented once, 12 twice, and 12 three times. The pairs were assigned randomly to these frequency categories. Pairs presented twice were first assigned positions randomly, once in each half of the study list. Pairs presented three times were then assigned, each pair occurring once in each third of the study list. Finally, the 48 pairs occurring once were assigned randomly to the remaining positions. Two buffer pairs were assigned at the beginning of the list, and two at the end so there was a total of 112 positions in the study list.

Test booklets. The test booklet consisted of three pages plus the cover sheet of instructions. There were 84 test items, 28 on each page. Each page included 16 items containing pairs that had been presented once for study, four that had been presented twice, and four that had been presented three times. In addition, there were four pairs on each page which had not been on the study list (new pairs). Half of the items in each class (excepting the new pairs) were tested by a forced-choice procedure, half by an absolute test. Thus, across the three pages, 24 pairs that had been presented once for study were paired with new pairs to produce 24 forced-choice items, and 24 pairs were presented alone for *yes–no* decisions. Pairs presented twice and three times for study were also equally divided between forced choice and absolute judgments.

[4] This pool of words was formed by Carl P. Duncan. Its use is greatly appreciated.

For absolute recognition, the words *yes* and *no* followed the pair with the *S* instructed to encircle the appropriate response. For the judgments of absolute frequency, a blank, rather than *yes* and *no*, followed the pair with the *S* requested to write in a number to indicate the frequency with which he thought the pair had been presented during the study list. In the forced-choice items the *S* encircled the correct pair (recognition) or the pair which had occurred most frequently during study (frequency judgments).

As noted earlier, new pairs were required for the forced-choice items. The 36 new pairs required (24 for the pairs presented once for study, six each for those presented twice and three times) had been formed randomly from the pool and were paired randomly with old pairs. On a single form only half the study pairs could be tested by a given method (forced choice or absolute judgment). Therefore, a second form was constructed in which the method of testing was reversed for the pairs. As a consequence, when considered across all *S*s, each study pair was tested by a forced-choice procedure (for recognition and for frequency information) and also by an absolute test (*yes* or *no* recognition, and judgments of absolute frequency).

On each page of the booklet the type of test item was randomized. Thus, *S* might have three successive items requiring a forced-choice decision, then an item requiring an absolute judgment, then an item for a forced-choice decision, then three items for an absolute judgment, and so on. On the forced-choice items the old and new pairs were randomly assigned to the left or right position. The instructions required that the *S* respond to all items on each page, guessing if necessary.

Procedure and Subjects. The data were collected by a group procedure. As the initial step, the booklets were distributed to the *S*s with instructions not to open the booklet. The experimenter informed the *S*s that they would be presented a long list of pairs of words and that they should try to learn the pairs as pairs. They were further told that some pairs would be presented more than once. Next, they were asked to study the instructions in order to understand how they would be tested. Any questions concerning the instructions were handled individually. The list was then presented for study. After the last slide was presented the *S*s were asked to review the instructions (if necessary) and then proceed to the first test page. The testing was unpaced.

The *S*s were tested in groups of varying size, but within each group both frequency-judging instructions and recognition instructions were represented as well as were both forms. The booklets were randomized within blocks of four (two forms and two types of instructions) before they were assigned to the *S*s. A total of 152 *S*s was completed, 76 having had the recognition instructions and 76 the instructions for frequency judgments. Within each group there were 38 for each form. Since the forms did not differ in any of the analyses, no further mention will be made of this balancing variable.

Results

Pairs presented two and three times were included merely to make the frequency-judging task a reasonable one, particularly for the absolute judgments. The errors made on these pairs were few in number. The basic interest was for the pairs presented once for study. The results for the forced-choice tests will be considered first.

Forced-choice tests. There were 24 items each of which consisted of a pair presented once and a pair which was new. The *S*s were required in one case to encircle the pair which had been presented for study, and in the other case, to encircle the most frequently presented pair. For the 76 *S*s given recognition instructions the mean number of errors was 2.59 (10.8%), and for the 76 subjects given instructions to choose the most frequently presented pair the mean number of errors was 3.3[illegible] (13.8%). The difference between the two means was not reliable, $t\ (150) = 1.73$, $p > .05$. The discriminability or sensitivity, therefore, may be judged equivalent for the two groups. This result is the same as found previously with single words (Underwood, 1972). In the previous study it was found that the errors fell on the same items under both sets of instructions. In the present study the number of errors made on each of the 48 pairs presented once for study was determined for each type of instruction. The product-moment correlation between the two arrays was .43, $p < .01$.

Absolute tests. For these tests, each *S* was given 24 pairs presented once and 12 new pairs. In one case *yes–no* decisions were requested, in the other, absolute judgments of frequency. The data will be viewed in a number of ways.

The frequency estimates may be handled in exactly the same way as are the recognition decisions. If *S* indicated that a pair not presented for study had a frequency of one or greater, it would be equivalent to a false alarm for a *yes–no* recognition decision. Similarly, if *S* assigned a zero to a pair that had been presented once, it would be equivalent to a miss in recognition. Throughout the studies to be reported, it has been found that the sum of the misses

and false alarms is a simple and meaningful measure of discriminability or sensitivity. This measure and d' correlate highly. In the present experiment the correlation for the 76 *S*s given the recognition test was −.86, and for those making absolute frequency judgments, −.91. These are somewhat lower values than those to be reported in later studies. In view of these relationships, the sum of the misses and false alarms (M + FA) has been used as the basic measure of discriminability.

The mean M + FA for absolute recognition was 7.66, that for frequency judgments, 6.∴. These means do not differ statistically (t = 1.20). However, these values alone do not adequately reflect the outcome for the absolute tests. The full picture can be seen in Figure 1. This figure implicates criterion differences between those *S*s who made recognition decisions and those who made frequency decisions. The *S*s in making recognition decisions infrequently made false alarms but correspondingly had many misses. The *S*s in making frequency judgments behaved in the opposite manner; they made many false alarms (assigning a new pair a value of one or more) but had few misses.

The difference in the criterion set by the two groups is reflected directly by the interaction (F = 69.66) as seen in Figure 1. An analysis of variance of data such as Figure 1 is based upon allows conclusions about both differences in sensitivity or discriminability and differences in the criterion or response bias. If there is a significant main effect between recognition and frequency judgments, differences in sensitivity would be indicated. A main effect of input frequency would simply indicate the overall balance between the tendency to produce false alarms and the tendency to produce misses. A significant statistical interaction resulting from a convergence of the lines between input frequencies zero and one (whether a crossover or not) indicates a criterion difference.

The above technique for determining sensitivity and criterion differences seems straightforward. However, some comments should be made about other measures used

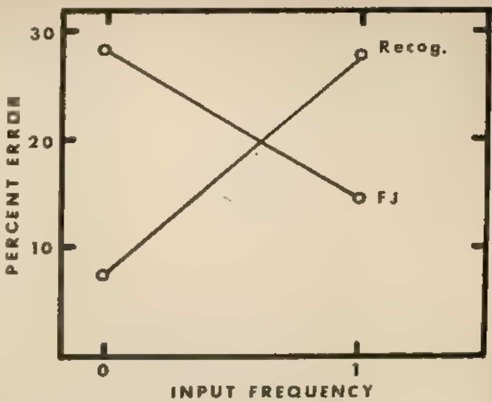


FIGURE 1. Errors on pairs of words as related to input frequency. The test included both frequency judgments (FJ) and recognition (Recog).

to reflect criterion differences. *Beta*, the measure of the criterion setting used in signal-detection analysis, is the ratio of the height of the ordinate of the normal curve for the proportion of hits and the ordinate height for the proportion of false alarms. The *beta* distributions were severely skewed for the present data. Ignoring this skewness, the mean *beta* for recognition was 4.95, and for frequency judgments, 2.44 (t = 3.59).

Another measure which could be used to reflect differences in the criterion is M − FA/M + FA. Thus, if one *S* had two misses and eight false alarms, and another *S* had eight misses and two false alarms, sensitivity is equivalent but the criterion differs for the two (−.60 for the first, and +.60 for the second). Scores may vary from −1 to a +1. For the present data the mean values were .72 for recognition and .08 for frequency judgments (t = 7.14). This measure is curvilinearly related to *beta*. If zero false alarms and zero misses are assumed to be one miss or one false alarm, the rank-order correlation between the two measures (*beta* and M − FA/M + FA) is essentially perfect. The logic of changing zeros to a positive value (the value could be less than one, e.g., .1) is to avoid too many extreme scores (−1 and +1) which result when zero scores are entered. This simulates the situation for *beta* where no ordinate value is zero.

The values for M + FA are essentially equivalent for the two groups. Therefore, the criterion difference resulting from the formula M − FA/M + FA must come from the M − FA values. In effect, this is simply another way of calculating an interaction. For the present work, therefore, any criterion difference between conditions seems given most directly by the interaction of the analysis of variance between input frequencies and type of test. However one analyzes the data, it is clear that there were differences in the criterion setting for *Ss* given the recognition task and those given the frequency-judging task. As was the case with forced-choice tests, sensitivity did not differ for the two groups on the absolute tests.

Item correlations. The number of misses for each of the 48 pairs presented once was determined for recognition and for frequency judgments on the absolute tests. The correlation between the two distributions was .63 ($p < .01$). This indicates a substantial relationship between phenomenal frequency and recognition decisions. This relationship emerges in another analysis. The mean frequency judgments for the 48 pairs presented once were correlated with the number of misses for recognition. The value was −.50 ($p < .01$).

Correlations by Subjects. A given *S* made both absolute and comparative frequency judgments, or both absolute and comparative recognition decisions. The relationship between the two types of judgments may be examined. For frequency judgments the correlation for the 76 *Ss* between M + FA for 24 items and number of errors on the 24 forced-choice items was .65. For recognition, the corresponding value was .59. Both values are reliable statistically ($p < .01$).

Each *S* under frequency-judging instructions made absolute frequency estimates for 24 pairs which had been presented once for study. For each *S* the mean and standard deviation of these judgments was determined and correlated with the number of errors made on the forced-choice frequency judgments. The correlation between the mean frequency judgments and number of forced-choice err[illegible] was −.09 but that between the stand[illegible] deviation and number of forced-choi[illegible] errors was .44 ($p < .01$). Thus, the l[illegible] the variability of phenomenal freque[illegible] for items with a constant input freque[illegible] the better the discriminability on [illegible]ed-choice decisions.

Discussion

The discussion of the finding[illegible] be brief at this point. Overall, the resul[illegible] showed that frequency judgments and recogn[illegible] decisions for pairs of words have much in [illegible]mon, and these results could be interprete[illegible] consistent with the idea that phenomenal [illegible]uency differences play a major role in re[illegible]ition performance. Alternative interpr[illegible]tions are undoubtedly possible, of course, [illegible]t such interpretations must include an a[illegible]ounting of the commonality in errors produce[illegible]by recognition and by frequency-judging in[illegible]uctions.

The finding that there was a diff[illegible]ence in the criterion set (on the average) for recognition judgments and for frequency jud[illegible]ments was quite unexpected. It was therefore decided to undertake a further study, with somewhat changed procedures, to see if the finding could be replicated.

Experiment II

Method

In this experiment all *Ss* were given exactly the same study list as was given in Experiment I. The two changes, which occurred on the test, will be described.

In Experiment I, each *S* made both absolute and comparative judgments. In Experiment II, e[illegible] *S* made a single type of judgment on all items. Therefore, there were four groups of *Ss*. One group made only absolute frequency judgments, [illegible] made only forced-choice frequency judgments. A third group made *yes–no* recognition decisions, and a fourth made forced-choice recognition decisions. Of primary interest were the judgments [illegible] on the 48 pairs presented once for study.

It will be remembered that in Experiment I only 12 new pairs were used for *yes–no* recognition and for absolute frequency judgments. In Experi[illegible] II, 48 new pairs were included. Since there [illegible] also 48 pairs presented once, the number of [illegible] rights and wrongs was more nearly equali[illegible] Experiment II than was true in Experiment [illegible] These 48 new pairs also became the new pairs on t[illegible] forced-choice tests.

The booklet for the forced-choice tests (recognition and frequency judgments) consisted of the

instruction sh... and two test sheets with 30 items on each for ... of 60 test items. These 60 items consisted ... 18 pairs presented once (each accompa... new pair) and six items in which the stud... had been presented twice and six in which i... been presented three times. Thus, only half ... items presented twice and half those presente... three times were tested. The purpose in this wa... keep the total number of pairs to which the sub...t was exposed on the test equivalent for the f... choice and for the absolute tests.

The test booklet for the absolute judgments was mad... of the instruction sheet and three test pa... items to a page, for a total of 120. These con...ed of 48 new pairs, 48 pairs presented once, 12 ...ated twice, and 12 presented three times.

...ther details of the procedure were exactly the sa... for Experiment I. The *Ss* studied the in...on sheet which described how they would be te...d, after which the list was presented for study at ... 5-sec. rate. Again, the test was unpaced. The four instructional conditions were randomized in ...ocks of four so that for each session each was about equally represented. Sessions were continued until 36 *Ss* had completed the test for each instructional condition.

Results

Forced choice. The mean number of ...rors made by the 36 *Ss* on the 48 forced-...oice items under recognition instructions ...s 5.42 (11.3%), and under instructions to choose the most frequent pair, 4.92 (10.1%). The t was .60. The correlation of errors by items was .54 ($p < .01$). These results confirm those of Experiment I.

Absolute tests. The mean frequency judgments for the pairs presented 0, 1, 2, and 3 times may first be noted. The means in corresponding order were .18, .91, 1.74, and 2.49. The relationship is linear. Again, however, in the analyses to follow, no attention will be given to the pairs presented more than once since the errors on these pairs were few in number. As was done for Experiment I, false alarms and misses were determined for the groups making absolute judgments. The results are plotted in Figure 2. The interaction, although less severe than in Experiment I, was still present, $F\ (1, 70) = 5.99$, $p < .05$. There was no appreciable effect of type of instruction ($F = 2.18$), indicating that discriminability was equivalent for both frequency decisions and recognition decisions. There was a main effect of input frequency

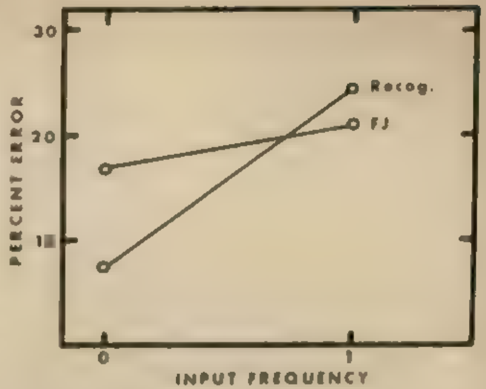


FIGURE 2. Further evidence on the errors made on pairs of words as related to input frequency for frequency judgments and recognition.

($F = 17.54$) indicating more misses than false alarms when combined across both groups.

The sum of the false alarms and misses correlated $-.93$ with d' for the frequency judgments, and $-.94$ for the recognition data. The difference in the criterion set by the two groups, as given in the interaction in Figure 2, was not reflected in *beta* differences, the F being less than one.

Item correlations. The number of misses was determined for each of the 48 pairs presented once for study. This was done separately for the 36 *Ss* making absolute frequency judgments (a miss being an assignment of zero) and for the 36 *Ss* making *yes–no* recognition decisions. The correlation between the two arrays was .55 ($p < .01$). The correlation for the number of false alarms on the 48 new items for the two groups was .42 ($p < .01$).

If frequency information is involved in recognition decisions, it should be possible to predict forced-choice recognition and frequency judgments from the absolute frequency judgments. In the forced-choice tests, two pairs were involved for each item, one old (presented once) and one new. For each of these pairs a mean frequency judgment was available from the group making the absolute frequency estimates. It would seem that the greater the phenomenal frequency difference between the two pairs in the forced-choice item, the

fewer the number of errors which should be made on that item. However, the variability of the frequency judgments should also be considered. For whatever reason, items having the same input frequency differ in the variability of the frequency estimates when viewed across the 36 estimates from the 36 *S*s. It was shown in Experiment I that *S*s having the lower variability of frequency estimates with constant input frequency did better on the forced-choice items than did the *S*s with the higher variability. The same outcome may apply to item variability when calculated across *S*s. Therefore, it would seem that the clearest prediction would be that the greater the mean difference in phenomenal frequency between an old and new pair in a forced choice item, adjusted for differences in the variability of frequency estimates for those pairs, the better should be the performance on the forced-choice test.

The mean frequency difference for each old and new pair appearing in the forced-choice items was divided by the square root of the sums of the variances of the frequency estimates for the two pairs. A high value should indicate good discriminability between the two pairs. The correlation between these values and the number of errors on the forced-choice recognition test was −.61. For the forced-choice frequency judgments, the correlation was −.79. Clearly, the phenomenal frequency of items presented once, and the variability of the estimates reflecting phenomenal frequency, predict performance on the forced-choice tests with some accuracy.

As one further bit of evidence on the relationship between frequency information and recognition performance, the absolute judgments made on the 48 pairs occurring once were combined across Experiments I and II. Therefore, there were 74 *S*s who had made absolute frequency judgments and 74 *S*s who had made *yes–no* recognition decisions. The apparent frequency of a pair was given by the mean of the frequency judgments, and for each pair a standard deviation was available. It would be expected that the higher the mean apparent frequency and the lower [illegible] variability the better should be the p[illegible]ormance on recognition. The coeffici[illegible]f variation (M/σ) was used to derive a[illegible]gle measure for each pair. A high co[illegible]ent should indicate good discriminabil[illegible] These 48 values were correlated with [illegible] number of errors on *yes–no* recogniti[illegible]or the 48 pairs. The product-mom[illegible]correlation was −.70.

Discussion

Both experiments have sh[illegible] that discriminability or sensitivity for [illegible] of words was equivalent when *S*s were [illegible]d to make recognition decisions and judgments of frequency. This outcome held for both forced-choice items and for absolute judgments. Furthermore, the particular ite[illegible] on which errors were made in recognition [illegible]nded to be the same items on which the errors were made when judging frequency and frequency differences. Mean frequency judg[illegible] for pairs combined with the variability of those judgments, predicted errors in rec[illegible]ition performance on both *yes–no* and forced-choice tests. Viewed as a whole, the results were entirely consistent with the theory which specifies that frequency information is the major attribute of memory involved in pair recognition. In short, these two studies indicate that when pairs of words constitute the unit of analysis of recognition the outcome is much the same as when the individual word is the unit.

An unexpected finding in both experiments was that *S*s making absolute frequency judgments set a more lenient criterion than did those *S*s making recognition decisions. The reason for this difference is not known. An examination of the changes in performance across pages of the tests showed only very slight decrements under any of the conditions of testing, and the differences in the criterion were present initially and remained relatively constant across pages. Perhaps background frequency influenced the frequency judgments and not the recognition decisions although, as noted in the introduction, other data would argue against this possibility. By way [illegible] looking ahead it may be noted that a later experiment will show that criterion differences were found to be present for different class[illegible] of pairs of words for the same *S*. For the time being, therefore, the problem concerning th[illegible]

reason for the criterion differences will be set aside.

The two studies reported have established nothing concerning the role of the association in recognition. They have merely shown that frequency theory does not break down when pairs of words are used, pairs which presumably have some associative relationship as a consequence of the study trial. To assess in a more analytical way the role of associations in recognition requires different types of studies. The immediately following experiments represent one of these different types.

Experiment III

Assume the *S* is presented a series of pairs of words under instructions to associate the words in each pair. Let three of the pairs be identified as A–B, C–D, and E–F. On the test of recognition he is presented A–B (intact pair) and C–F (broken pair). His instructions are to respond *yes* to all pairs providing both words in the pair had been presented for study whether paired during study or not. If the number of misses is greater for the broken pairs than for the intact pairs, the association would seem to be clearly implicated in recognition performance. If the misses are equivalent, it would be concluded that the association is irrelevant for recognition performance. The purpose of Experiment III was to make this test.

In addition to the two types of test items noted above, two others were included in order to have pairs for which the correct response was *no*. One of these types consisted of one old and one new word, and the other type consisted of new pairs.

Method

Materials. The pairs of words were from the same pool as used in the first two studies. The *S* was presented 54 pairs for study. Of these, 18 were tested as intact pairs, that is, both words were old and paired as they were for study (O–O). A further 18 were re-paired words or broken pairs (O–Ob). One word from each of the remaining 18 pairs was used in pairs in which a new word constituted the other member of the pair (O–N). Finally, the test included 18 new pairs (N–N pairs).

For purposes of balancing conditions against possible pair differences, each pair was used in all conditions equally often. Originally, 72 pairs were assigned randomly to four subgroups of 18 pairs each. Across four forms each subgroup was used once for each of the four test types as described above. Thus, there were four different lists presented for study and four different test forms. The only additional words required were the 18 used as new words for the O–N pairs. These 18 words served the same function across all forms although, of course, the old words differed for each form. The O–Ob test pairs never consisted of an interchange of the members of two pairs. In the presentation list, each type of test item (except N–N of course) was represented in each successive block of three pairs. On the test, item types were randomized within blocks and particular items within blocks were random with respect to presentation position during study.

One further variable was introduced in the test. Half of the O–O pairs were tested in the order presented for study (A–B), and half were tested in reverse order (B–A). This was also true for the O–Ob pairs and for the O–N pairs. The latter became N–O pairs when reversed. The particular nine pairs to be reversed were determined randomly.

Procedure and Subjects. The pairs were presented for study at a 5-sec. rate by a slide projector. The instructions emphasized the learning of pairs as pairs and, although the *Ss* knew their memories would be tested, the nature of the test was not specified before presenting the list. After the study of the pairs the *Ss* were given a two-page booklet, 36 pairs to a page, with the words *yes* and *no* after each pair. They were instructed carefully that *yes* was the appropriate decision if both words had been [illegible] study or not. The instructions further empha[illegible] that if one or both of the words in a test pair had not been presented for study, the appropriate response was *no*. According to these rules there were 36 pairs for which the correct response was *yes*, and 36 for which the correct response was *no*, although the *Ss* were not told this.

Subjects were tested in small groups until 18 were completed for each of the four forms. Forms did not prove to differ statistically on the test, and they did not interact with item types. The results will be considered for the 72 *Ss* combined.

Results

The overall results are shown in Figure 3, where percent errors is used as the recognition measure. An error would be a false alarm for N–N and O–N, and a miss for O–O and O–Ob. It is to be noted that there were about 7% more misses on O–Ob pairs than for O–O pairs. The overall analysis showed a significant effect of item type ($F = 37.58$). A standard error of the mean difference derived from the error term indicated that any difference as large as

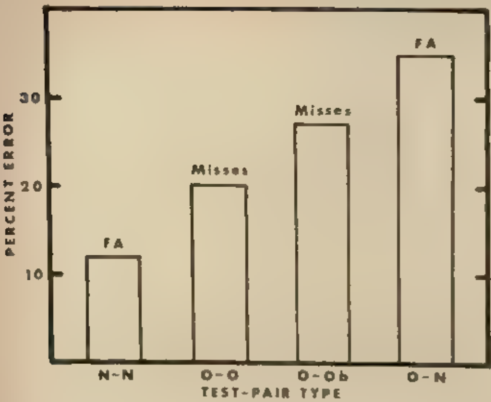


FIGURE 3. Misses and false alarms (FA) for four types of test pairs. (Abbreviations: N refers to a new word, O to an old word, and b to a broken pair.)

4.3% was reliable at $p = .01$. All differences between item types were larger than 4.3%. The fact, therefore, that type O–O results in fewer errors than type O–Ob may be taken as evidence that the association between the two words in a pair played a role in recognition.

The misses on the O–O pairs were more frequent than were the false alarms on the N–N pairs. The M + FA measure of sensitivity for these pairs correlated with d' across the four forms at the following levels: −.93, −.94, −.91, −.95.

The O–N pairs produced nearly 35% false alarms. A satisfactory accounting of this result was not found. Several subsidiary analyses were done to search for leads. The 72 *S*s were divided into two subgroups based upon the total errors made on all item types. The relative relationship among the four item types were exactly the same for these two subgroups of good and poor *S*s. There were two pages to the test form, and the *S*s completed the first page before moving to the second. It seemed possible that testing effects might have differed for item types. However, an analysis showed no difference in performance on the two pages and no interaction between pages and item types, both *F*s being less than 1.

It will be remembered that the position of the words in a pair during study was reversed for half the pairs on the test. This change had essentially no effect on the O–O pairs, the nonreversed pairs producing 20.7% errors, the reversed pairs, 19.6%. For the O–Ob pairs the corresponding values were 25.8% and 27.9%. But, for the O–N pairs the false alarms were 38.0%, and when reversed, (N–O) 31.6%. This is to say that when the old word was the first member of the pair, errors were more frequent than when the new word was the first member. However, even for the latter pairs, performance was poorer than was found on any of the three other item types.

It is difficult to see how criterion differences could exist for the various item types, since if the *S* could classify an item type he presumably could provide the correct response. Frequency theory might predict good performance on the O–N pairs on the grounds that the two words would represent a contrast in apparent frequency and this contrast would lead to a correct rejection. In one sense the results for the O–N pairs are in contradiction to the notion that an association facilitates recognition. The two words in the O–N pairs were not associated during study just as the O–Ob pairs were not associated. If the lack of an association leads to a *no* decision for the O–Ob pairs it should do the same for the O–N pairs, but in the latter case the decision would be correct. It is possible, of course, that this effect was present and kept the number of errors from being more frequent than was actually found. For the time being, the mechanisms responsible for the heavy error rate on the O–N pairs cannot be specified.

One other fact will be reported. When words are paired randomly the resulting pairs should show a distribution defined in terms of the ease with which the words within the pairs could be associated. There is a possibility that a pair which is easily associated when presented for study will also produce errors (false alarms) when it is used as a new pair on a test. This does not seem to be the case. In the present experiment, the 72 pairs were used both as old pairs (O–O) and as new pairs (N–N). The number of false alarms made on the

72 pairs when they were used as new pairs did not correlate with the number of misses which occurred when they were used as old pairs ($r = -.03$).

Experiment IV

The results of Experiment III were interpreted as demonstrating that an association between two words, developed on a single study trial, facilitated recognition of the pair. Or, as is sometimes said, associative context aided recognition performance. Experiment IV was another test of the role of associative context using a different approach, although one that is similar to those used by other investigators, particularly by DaPolito, Barker, and Wiant (1972). These investigators presented a triad of words for study. On the test, the *Ss* made recognition decisions on single words, but other items were sometimes present in the display when the decisions were being made. Essentially any change at test produced an increase in misses. The *Ss* were not instructed to associate the words in the triads during study, and since the triads were presented for only 1 sec., associative formation between the items would be minimal. Still, the results showed that when the order of the three items as presented for study was changed on the test, a decrement was observed, suggesting either that item-order information was important or that associations between the item and its position within the triad was a part of the memory. The method used by DaPolito et al. makes context (when defined as the presence of an item or items on which no recognition decisions are to be made) seemingly incidental. To some degree it would seem that *S* could ignore the context words on the test. The fact that context effects were found in spite of this seems to speak strongly that context, whether known to be associative or not, does influence recognition.

In the present experiment, *Ss* were presented pairs of words for study under instructions to associate the words in each pair. On the recognition test, either single words or pairs of words were presented. In the latter case the recognition decisions were made only on the second word in the pair. Thus, the first word in the pair need not have been involved at all in the recognition decision although the *Ss* were told that it might help them in reaching a decision on the second. Of course, the major variable was the nature of the first word in the pairs.

Method

A description of the seven classes or types of test items will be given first, expanding the symbol system used in describing the pairs in Experiment III. An old item is designated O, a new item, N, and (*a*) the test word was always the second word in a pair on the test, and (*b*) a word on the test always occupied the same position in a pair (first or second member) as it held during the study trial. The seven types of items were as follows.

- Type O: old second word tested alone
- Type N: new word tested alone
- Type O–O: intact pair
- Type N–N: new pair
- Type N–O: old second word with new first word
- Type O–N: new second word with old first word
- Type O–Ob: two old words from different study pairs

Materials. The words in the pairs were all four-letter monosyllabic words. They were drawn from a pool of 312 such words which constituted a random sample from the Thorndike Lorge tables. The change in the class of words (from the class used in the first three experiments) was necessitated by the fact that other experiments being conducted in the department were using the words from the earlier experiments.

Study and test lists. Initially, 96 words were drawn randomly from the pool and formed into 48 pairs randomly. These 48 pairs constituted the study list, hence the old pairs, for all *Ss*. A second sample of 84 words was drawn randomly and these 84 words were always used as new words on the test form.

For balancing purposes, the 48 study pairs were arbitrarily divided into four blocks of 12 pairs each. Across four test forms, each block of 12 pairs was used once for each of the four item types on the test involving old words in the second position (O, O–O, N–O, O–Ob). This means, therefore, that on a test form there were 12 items representing each of the four types. The O–Ob pairs were never constructed by interchanging the members of two pairs.

The assignment of new items to particular functions on the test forms was done randomly, with a different random assignment for each form. There were 12 Type O–N test pairs, the O words being the first words in the study pairs which on the test became Type O. There were also 12 Type N pairs

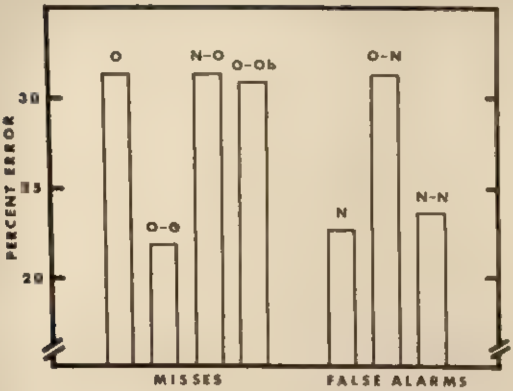


FIGURE 4. Misses and false alarms for various types of test items. (For pairs, the recognition decision was made only on the second word in the pair.)

on the test, but 24 pairs were used to represent Type N–N. This made 48 true *yes* items on the test and 48 true *no* items. On the test, the seven item types were block randomized, and test position, relative to input position, was orthogonal by halves. There were 48 items on each page of the two-page booklet.

The single items on the test always occupied the nominal second position of a pair. For example, the first four items on one of the test forms were as follows.

1.	spry–wile	YES	NO
2.	whey–ramp	YES	NO
3.	jest	YES	NO
4.	came–dell	YES	NO

Procedure and Subjects. The 48 study pairs, preceded and followed by two filler pairs, were presented at a 5-sec. rate. Prior to presenting the pairs, the *S*s were told only that pairs would be presented and that they should try to associate the words in each pair in preparation for a memory test. After the last pair had been presented, the booklets were distributed. The *S*s were run in groups of varying size but in any group the four test forms were assigned randomly to the *S*s. The instructions for the test involved five key points: (*a*) for a single word on the test form, the *yes–no* decision was whether or not the word had been shown as a member of a pair; (*b*) for a pair of words, the decision was to be made only for the second word, i.e., had it or had it not been presented as a second word in a pair on the study list; (*c*) the first word in a pair might aid in making the decision on the second; (*d*) an old word always occupied the same position on the test form as it had in the pair in which it appeared on the study list; and (*e*) all items were to be completed, guessing if necessary. The test was unpaced.

Groups of *S*s were tested until 68 *S*s had been completed, 17 for each of the four forms.

Results

The results, shown in Figure 4, are quite unambiguous. There were only two levels of errors, differing about 8–10%, and these two levels were associated with change of context and with context constancy. For the misses, omission of the first word (O), including a new first word (N–O), or repairing old words (O–Ob) all resulted in about the same increase in the number of errors when no change (O–O) is used as a reference. False alarms were correspondingly influenced when an old word was used as a first word (O–N). A number of different ways of viewing the differences statistically all produced the same conclusions. Perhaps the most direct way is to ask about the differences among the four types for the misses, F (3, [illegible]) = 7.87, $p < .01$, and among the three types for the false alarms, F (2, 134) = [illegible], $p < .01$. In both analyses, the deviation of one condition from the others is largely responsible for the conclusion that reliable differences were present.

It will be remembered that in Experiment III, Types N–O and O–N produced more errors than did Type O–Ob. In the present data, where *S* was asked to make decisions on a single word rather than on a pair as was true in Experiment III, the numbers of errors for these types were essentially equivalent. Also, in Experiment III the misses on Type O–O exceeded the false alarms on Type N–N, which was not true in the present experiment. This is to say that the two techniques (judging single words in pairs versus judging pairs) produced differences in detail, but both lead to the same conclusion. The presence on the test of the full associative context present during study facilitated recognition. Any changes made in the context caused an increase in misses.

EXPERIMENT V

Experiments III and IV gave evidence that an intact association between two words in a pair is a positive factor in the recognition of that pair. The present experiment represents a further test of the

role of an association in recognition, a test which makes use of paired-associate learning.

It was mentioned earlier that when words are paired randomly one would anticipate that some of these pairings would consist of words which would be easy to associate and others which would be difficult to associate, with the bulk of the pairs lying in between the extremes. If an association between two words influences the recognition of the pair, it might be reasoned that old pairs on which few errors were made developed stronger associations than did pairs on which many errors were made. To put this another way, during a constant period of study, the strength of the acquired association would differ for different pairs. If the strength of the association is involved in recognition, it must be predicted that pairs on which few recognition errors were made would be learned more rapidly as paired associates than would pairs on which many errors were made. Experiment V was a test of this proposition.

Method

Materials. From among the 48 pairs presented once for study in Experiment I, and tested by a *yes–no* procedure, 10 pairs on which few recognition errors were made were selected along with 10 pairs on which the most recognition errors were made. As paired-associate lists, these will be called the Easy List and the Hard List. Both lists are shown in Table 1. Although these lists were selected by using the data of Experiment I, the same pairs occurred under much the same conditions in Experiment II. Therefore, the errors made in Experiment II are also shown. The reliability of pair difficulty is by no means perfect, particularly for the pairs in the Hard List, although some regression would be anticipated. Nevertheless, the mean errors were substantially different for the two lists even when gauged by the errors made in Experiment II.

TABLE 1

EASY AND HARD LISTS USED IN THE PAIRED-ASSOCIATE (PA) EXPERIMENT WITH OTHER RELEVANT INFORMATION ABOUT THE PAIRS

List	Recognition errors		Frequency judgments		Correct in PA
	Experiment I	Experiment II	*M*	*M*/σ	
Easy					
nymph-prune	1	4	.95	1.53	67
bacon-shrub	3	6	1.11	1.28	58
mound-alert	3	15	1.24	1.29	65
onion-ivory	4	4	.93	1.37	75
daddy-waver	5	3	.86	1.30	77
broom-folly	5	5	1.26	1.34	61
laden-cargo	5	8	1.09	1.33	55
drank-voter	6	6	.99	1.43	76
exalt-baron	6	3	1.05	1.52	62
wedge-canon	7	6	.97	1.26	53
M	4.5	6.0	1.05	1.37	64.9
Hard					
hatch-focus	25	21	.80	.84	45
marsh-brute	20	13	.76	.96	23
award-birch	17	6	.94	1.07	54
dwelt-forge	17	4	.99	1.27	19
harsh-inner	16	17	.81	1.01	44
lance-chime	16	21	.73	.97	44
risen-stray	16	18	.86	.97	43
punch-exile	16	8	1.08	1.33	32
thief-drill	15	10	.78	.96	66
flush-abide	15	11	.89	1.16	22
M	17.3	12.9	.86	1.05	39.2

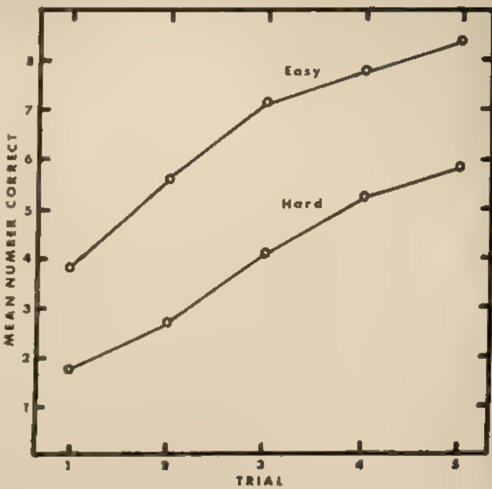


FIGURE 5. Acquisition of paired-associate lists of 10 pairs each, identified as easy and hard by the number of recognition errors made on the pairs.

Procedure and Subjects. The lists were presented for one study trial and five anticipation trials at a 2:2-sec. rate on a memory drum. There were 20 *Ss*. Half learned the Easy List followed by the Hard List, and half learned in the reverse order.

Results

As may be seen in Figure 5, the list made up of pairs defined as easy by recognition errors was much easier to learn than the list made up of pairs on which many recognition errors were made. This was true regardless of the order in which the lists were learned and the overall difference across the five trials was highly reliable ($F = 28.09$).

The mean number correct for each pair, summed across the five trials, is shown in Table 1. Only two pairs from the so-called Hard List fall within the range of number

of correct responses as shown for the pairs in the Easy List. Other information is included in Table 1. In Experiment I, 38 *S*s made absolute frequency judgments of these pairs, and in Experiment II, 36 *S*s did the same. The mean apparent frequency as derived from these 74 *S*s for each pair is shown in Table 1. The difference in the mean frequency judgments for the two lists is significant, $t\ (18) = 3.69$, $p < .01$. Thus, the pairs in the Easy List had higher apparent frequency than did those in the Hard List.

As shown earlier, prediction of recognition errors from absolute judgments of frequency was quite good when means and variability were considered simultaneously by using the coefficient of variation (M/σ). These values also appear in Table 1 and it is to be noted that the two lists differ markedly on this measure. Rank-order correlations were calculated across the 20 pairs. The M/σ values correlated .86 with recognition errors and .56 with the number of correct anticipations in paired-associate learning. The correlation between recognition errors and paired-associate learning was .72.

Discussion

The present results indicate that the ease of associating two words in a pair is positively related to correct recognition. These results support those of Experiments III and IV in showing that an association between the two words in a pair seems to be involved in recognition performance. Insofar as apparent frequency has been shown to predict recognition errors it is not surprising that apparent frequency also has some predictive value for paired-associate learning.

The evidence that an association between two words influences recognition must be kept in theoretical perspective. The major information used in recognition decisions may well be frequency information—apparent frequency of the pair as a unit and of the individual words, each as a unit. The difference between intact and broken pairs in Experiments III and IV was only 7–10%; performance did not fall to chance when associative context was removed. Thus, although the results of the experiments thus far indicate that an association plays a role in recognition, the magnitude of the role is not impressive when overall recognition performance is considered.

Experiment VI

Two of the basic findings from the preceding experiments led to Experiment VI. Experiments III and IV indicated that an association resulting from one presentation of a pair of words adds information which aids recognition performance. Experiment V showed that pairs on which few recognition errors were made were more rapidly acquired in a paired-associate list than were pairs on which many recognition errors were committed. These facts suggest that the associability of two words in a pair facilitated recognition performance, or, that facilitation of recognition of a pair by the association existing between the two words was directly related to the strength of that association. An obvious projection from these facts is that if very strongly associated pairs of words were tested for recognition, few if any errors would be observed. The present experiment deals with this possibility.

Many of the pairs used in Experiment VI were strongly associated as a consequence of cultural usage. Other investigators have approached the problem of associative context by the use of similar materials. Tulving and Thomson (1971) presented *S* strongly associated pairs of words as well as weakly associated pairs as indexed by published word-association norms. On the recognition test the *S* was asked to make a decision on each word of a pair independently. Based on the results of this procedure it would be concluded that the strength of an association had little influence on recognition. There were 12% misses on the strongly associated pairs, and 21% false alarms on new associated pairs. For weakly associated pairs the values were 15% and 16%. However, these data came from only a small part of an extensive procedure of testing which involved changing words in pairs, testing single words, adding new words to words which during study had been members of pairs, and so on. It is possible that the

TABLE 2

CRITICAL PAIRS OF WORDS: EXPERIMENTS VI–VIII

Antonyms	Conceptual associates	Parallel associates	Synonyms	Homonyms
day–night	canary–bird	cup–saucer	complete–entire	see–sea
bad–good	horse–animal	bed–sleep	spoken–verbal	pail–pale
lost–found	minnow–fish	bread–butter	hidden–concealed	plane–plain
hard–soft	maple–tree	income–money	empty–vacant	minor–miner
dirty–clean	python–snake	lamp–light	bungalow–house	peek–peak
boy–girl	waltz–dance	hammer–nail	capsule–pill	cent–sent
buy–sell	pansy–flower	needle–thread	chill–cold	weak–week
true–[illegible]	red–color	salt–pepper	correct–right	fare–fair
give–take	murder–crime	eight–nine	wicked–evil	sleigh–slay
high–[illegible]	water–liquid	scissors–cut	signature–name	tea–tee
king–queen	apple–fruit	spider–web	starved–hungry	stayed–staid
long–short	cancer–disease	table–chair	silent–quiet	sale–sail
slow–fast	bracelet–jewelery	ale–beer	grief–sorrow	wholly–holy
bottom–top	silk–cloth	army–navy	careful–cautious	mane–main
far–near	pliers–tool	candy–sweet	mad–angry	meat–meet
hate–love	emerald–gem	dock–boat	tiny–small	need–kneed
open–close	gnat–insect	hand–foot	central–middle	lode–load
rich–poor	east–direction	nurse–doctor	ancient–old	course–coarse
smooth–rough	uncle–relative	mail–letter	rural–country	prey–pray
under–over	oil–fuel	lock–key	boulder–rock	seen–scene

varied test context, and testing the words in a pair independently, may not have allowed differences as a function of strong and weak associative context to manifest themselves. Thomson (1972) used strongly associated pairs in one experiment (Experiment I) and pairs with no preexperimental associative relationship in another (Experiment IV). The *S*s made decisions on each word in a pair. The misses were 5% less for the strongly associated pairs than for those which were weakly associated, with the false alarms about equivalent for new pairs. These data, like the Tulving–Thomson data, suggest only a small facilitating effect of the culturally established associative relationship. Such findings seem odd if one is to assume that associative context is a powerful factor in recognition memory. Further investigation seemed necessary.

In the present experiment the *S* studied pairs of highly associated words and pairs of initially nonassociated words. Recognition decisions were always made for pairs as pairs.

Method

Materials. From a variety of sources, 100 pairs of words were brought together. As may be seen in Table 2, there were 20 pairs in each of five classes. Four of these classes were assumed to represent pairs with strong, culturally established, associative relationships. These four classes are the parallel associates, synonyms, antonyms, and conceptual associates. The fifth class, homonyms, represent a special class. There is no reason to believe that the two words in a homonym pair are associatively related in the sense that the words in pairs in the other classes are associated. The distinguishing property of the homonym pairs for the present study was that the acoustic-articulatory response to the two words in a pair would be equivalent. The consequence is that the frequency of this response would be double that for the words in the other types of pairs. It was believed that this property would yield information not only relevant for frequency theory, but also relevant to an understanding of the mechanisms of recognition in the associated pairs.

No argument is to be made for the purity of the pairs in the classes, nor that the pairs are necessarily representative of all possible pairs which might be placed in these classes. But, taken as a whole, each class is assumed to be distinctly different from each other class.

In addition to the 100 associated pairs, 100 nonassociated pairs were formed. In constructing these 100 nonassociated pairs, 100 additional words were chosen, these words varying in length, frequency, and form class. Some examples are: *assail, degree, lid, oath, worthless.* These words were paired randomly with one word from each of the 100 pairs listed in Table 2. Half the time (within each class), the first word in the pair as listed in Table 2 was retained, and it occupied the first position in the pair, the second position being held by a neutral word. For the other half of the pairs, the

TABLE 3

MEAN ERRORS (M + FA) FOR E PAIRS AND C PAIRS IN EACH OF THE FIVE CLASSES

Pair types	Parallel associates	Synonyms	Antonyms	Homonyms	Conceptual associates
E	1.14	1.15	1.70	1.42	.95
C	1.32	1.32	1.23	1.10	.83

Note. The standard deviations for these 10 distributions varied between .94 and 1.39. Abbreviations: M = misses; FA = false alarms; E = highly associated; C = nonassociated.

second word in the pair was retained, the first position being occupied by the neutral word. For example, using the pair *cup–saucer* in the list of parallel associates, the nonassociated pair became *cup–artist*, and using *bed–sleep*, the nonassociated pair became *utter–sleep*. Thus, the highly associated pairs (to be called E Pairs) and the nonassociated pairs (to be called C Pairs) always had one word in common.

Study lists and test forms. Each *S* was presented 50 pairs on the study list (plus two filler pairs at the beginning and two at the end). These 50 pairs were made up of five E Pairs and five C Pairs from each of the five classes. There were 100 pairs on the test forms consisting of the five E and C Pairs from each class presented for study plus five E and C Pairs from each class not presented for study. Of course, no word appeared in both an E Pair and a C Pair on the study list or test. For any given *S*, only half of the E Pairs and half of the C Pairs were used. Across four study lists and four test forms, each E Pair served as an old pair and as a new pair, and this was also true for each C Pair. This was accomplished by dividing the pairs in each class into four subgroups of C and E Pairs of five pairs each and rotating functions (old or new) across forms.

For the study list the 10 item types (five E Pairs and five C Pairs from each class) were block randomized across the 50 positions. The test forms consisted of 100 pairs, 50 on each of two pages, with the words *yes* and *no* appearing after each pair. The order of the pairs on the test forms was random, but with a different random order being used for each form.

Procedure and Subjects. Prior to presenting the list for study, the *S*s were fully informed concerning the nature of the test. They were told to try to associate the words in each pair on the study trial. The list was presented at a 5-sec. rate by a slide projector. After the list was presented the test booklets were distributed and the instructions for the test repeated. The *S*s were informed that they must make a decision for each pair, guessing if necessary.

The *S*s were run in small groups until 21 had completed each of the four forms. Since forms did not interact with the variables of experimental interest, the data have been summed across forms for the 84 *S*s.

Results

Data were available on 20 different types of pairs, each type represented by five pairs. These types were: old E Pairs in each of the five classes; new E Pairs from each class; old C Pairs from each class; and new C Pairs from each of the five classes. For the initial analysis, the sums of the misses and false alarms (M + FA) was determined for the C Pairs and for the E Pairs for each class. (It may be noted that d' correlated .98 with M + FA when summed across the five classes for the old and new E Pairs.) The means of the M + FA scores are shown in Table 3. It will be noted that for only two of the five classes (parallel associates and synonyms) are there fewer errors for the E Pairs than for the C Pairs. For the other three, the direction is reversed, although in none of the five classes is the difference between E and C Pairs of appreciable magnitude. Nevertheless, an analysis of variance showed the class was a significant source of variance, F (4, 332) = 7.87, $p < .01$, as was also the interaction between class and E-C Pairs, F (4, 332) = 4.18, $p < .01$, but the difference between E and C Pairs was not, F (1, 83) = 2.25, $p > .05$. These findings indicate that, overall, associated pairs are not better recognized than nonassociated pairs.

The measure of sensitivity, or discriminability, M + FA, has indicated that across the five classes the discriminability of pairs of highly associated words was not better than the discriminability for pairs with the minimal associative relationship that might have been established on a single study trial. The interaction between type of pair and class, however, indicated that discriminability differed as a function of class of pairs. An examination of the results for each class in more detail is indicated.

In presenting the results for the first two experiments, it was pointed out that criterion differences could be detected by a certain type of interaction when two conditions, both represented by old and new items, were plotted on the same graph. These requirements were met in the present

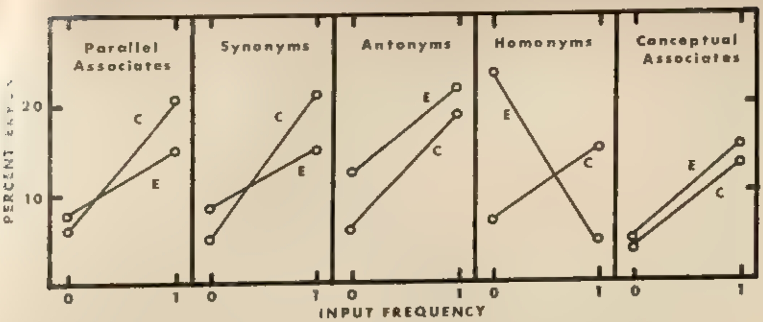


FIGURE 6. Recognition errors on experimental (E) and control (C) pairs of words as a function of input frequency and class or type of pair.

. for each class since there were old and new E Pairs and old and new C Pairs. The data are plotted in Figure 6 for each class, using input frequency (0 and 1) rather than old and new along the baseline.

The critical fact exhibited in Figure 6 is that *S* may maintain different criteria for different classes of pairs even when those pairs are intermingled on the same test form. The results for the homonyms show that *S* has a very lenient criterion for accepting pairs which were not in fact on the study list. The marked interaction between input frequency and E–C Pairs is evident. The results for each of the five classes of pairs have been analyzed separately. Each F is based on 1 and 82 *df*, and an F of 3.96 is required for the .05 level, 6.96 for the .01 level. Input frequency was significant for all classes, and except for the homonyms (which shows the smallest F, 9.51), more errors were made on old than on new pairs. The E–C difference was significant for only the antonyms ($F = 9.15$) and homonyms ($F = 4.02$). The interaction between the two variables was reliable for parallel associates ($F = 6.65$), synonyms (11.15), and, of course, for the homonyms ($F = 59.67$). The interaction for the parallel associates and for the synonyms cannot be interpreted as representing a criterion difference for the E and C Pairs, since the differences in errors are small in magnitude for new (zero frequency) items.

An examination of performance on the two test pages separately showed no systematic changes for any class of item. The differences in the criterion imposed for the different classes, therefore, was not a consequence of a gradual change or shift as the *S* made successive judgments.

It will be remembered that one word in each C Pair was tied to a word from a pair in one of the five classes. There is no reason to expect that these pairs should differ in number of errors across classes unless the words in the different classes differed in recognition difficulty in some way. Table 3 did suggest, however, that small differences were present among the C Pairs representing the different classes. Statistically, the M + FA values for the C Pairs for the five classes were reliable, F (4, 332) = 4.00, $p < .01$. The greatest number of errors occurred for the C Pairs derived from words making up parallel associates and synonyms, and the fewest from the conceptual associates.

Discussion

Four classes of natural associates have given no consistent evidence that the associative relationship facilitated recognition performance. Overall, control pairs, where the associative relationship would be minimal after a single study trial, produced no more errors than did the associated pairs. Even if it is presumed that the *Ss* rehearsed the C Pairs more than the E Pairs on the study trial, it does not seem possible that the associative strength of the C Pairs would even approach that of the E Pairs. It therefore becomes apparent that no simple conclusion about the role of associations in recognition is possible. This is not to say that an interpretation of the findings of Experiment VI as seen in Figure 6

is not possible. A tentative interpretation will be given later. Before doing this, however, it will be useful to look at the results of some additional experiments.

Experiment VII

In this experiment *S*s were again presented pairs of items from the five classes or types of pairs used in Experiment VI. On the test, only one word from each pair was tested, along with single words (control or new words) which had not been presented. The purpose of this procedure was not to test the effects of removal of associative context. Rather, the interest was centered in the relative recognition performance under these "impoverished" conditions for the five different classes. The interest was in particular directed toward the homonyms. It had seemed clear from Figure 6 that a lenient criterion was involved in the judgments on the homonym pairs. This alone could result in few misses on the E Pairs. But still another factor would also keep the number of misses to a minimum. As discussed earlier, a homonym pair should double the frequency of the acoustic–articulatory response as compared with that for any single word in the pairs of the other classes. Insofar as frequency information enters into the recognition decisions for pairs, the misses should be less for the homonym pairs than for the other types. Now, when a single word from a homonym pair (as presented for study) is presented for a *yes–no* test, the acoustic–articulatory frequency should be the major basis for the decision. In fact, it should be of little consequence whether the pair or the single word is presented if frequency is the dominant attribute. Number of misses should be small in both cases. Consider next the case of a single new word being tested, a new word which is a nominal member of a homonym pair but a pair which was not presented for study. There seems to be no reason to expect that the *S* would perceive this word as having a homonym. Therefore, there should be no criterion problem that is specific to homonyms. Or, in terms of formal predictions for the experiment, new items (from nominal homonym pairs) should not produce more false alarms than single new words representing the other classes. And, as discussed above, misses should be fewer on the old words from homonym pairs than on the old words from the other pairs.

Method

Study lists and test booklets. The *S* was presented 50 pairs for study, 10 pairs from each of the five classes. There were also two filler pairs at the beginning and end. On the test, one word from each of the 50 pairs was printed in a booklet along with 50 new words, representing one word from each of the 50 pairs remaining in the five classes. For the 50 pairs presented for study, half the time (within each class) the first word in the pair was tested and half the time the second word in the pair was tested. By using two study lists, and a single test form, each word served once as an old word and once as a new word. The study list was block randomized so that each type of pair occurred once in each block of five trials. On the test form (two pages) the 10 item types (five old and five new from each class) were block randomized and were random with respect to input order.

Procedure and Subjects. Prior to presenting the list of pairs for study, the *S*s were informed that the study list consisted of a series of pairs, and further, that on the test there would be single words, some from the pairs presented and some from pairs not presented. They would be required to make a *yes–no* decision on each word. The list was presented by a slide projector at a 5-sec. rate. Immediately following the presentation of the last slide, the booklets were distributed, and the instructions for the test repeated. As was true in all of the previous studies, the test was unpaced and the *S* was required to complete all items on the first page before going to the second. Thirty-four *S*s completed each form for a total of 68.

Results

Figure 7 shows the basic results, plotted in the same manner as Figure 6. Two facts are to be noted. First, errors on the antonyms (both misses and false alarms) were considerably higher than for the other types. Second, the fewest misses were made on the homonym pairs. The statistical analysis shows that class or type of item, based on M + FA is reliable, F (4, 268) = 18.72, $p < .01$. More errors were made on old words ($f = 1$) than on new words ($f = 0$), F (1, 67) = 6.18, $p < .05$, and the interaction between frequency and type was also reliable, F (4, 268) = 6.52,

$p < .01$. Sources of these effects are quite apparent in Figure 7.

The number of misses on the homonyms was slightly under 14%. The class of items having the next fewest misses was conceptual associates 19%. A direct-difference test showed this difference to be reliable, $t(67) = 2.84$, $p < .01$. This may be interpreted to mean that input frequency was greater on the homonym pairs than in the other types. The fact that there were only slightly more false alarms (yes to new items) for the homonyms than for the corresponding items in three of the other classes is interpreted to mean that a criterion difference for the classes was of little consequence in this study.

Recognition of antonyms was clearly inferior to the recognition of the words in the other classes. There was a suggestion of this possibility in the previous experiment for the E Pairs (see Figure 6). A part of the difference seen in Figure 7 can be attributed to heavy testing effects for the antonyms. This is to say, performance decreased in accuracy from page 1 of the test booklet through page 2. This was quantified for each class of items by using M + FA, with 10 items on each page. The errors increased for the parallel associates (15.3% to 19.3%), for the synonyms (16.9% to 22.2%), and for the antonyms (20.7% to 31.5%). All three of these changes were reliable beyond the .05 confidence level. The homonyms showed an unreliable increase (14.9% to 16.5%) and the conceptual associates showed a slight decrease (17.6% to 16.2%). The reasons for these differential testing effects in this experiment are not clear. It should be noted, however, that as pointed out earlier, there were no systematic changes in test performance over pages for Experiment VI where pairs were tested.

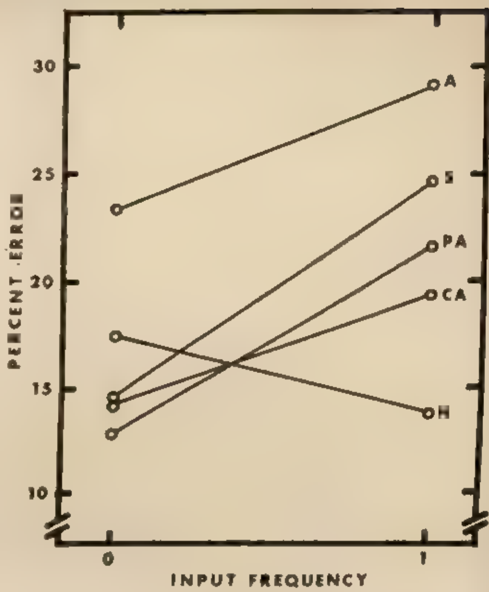


FIGURE 7. Recognition errors on single words following pair presentation for antonyms (A), synonyms (S), parallel associates (PA), conceptual associates (CA), and homonyms (H).

EXPERIMENT VIII

This experiment will be briefly reported. It had as its major purpose the examination of the possibility that some peculiarity of the words used in the homonym pairs led to the small number of misses in Experiments VI and VII. It is, perhaps, becoming apparent that some theoretical importance is to be attached to the findings reported thus far for the homonyms. The test made in the present experiment was simply to present single words from the pairs in each class for study and then test these single words for recognition along with new words from the classes. The critical question is whether or not the outcome will be the same for single words from the homonym pairs as for the single words from the other classes, particularly the classes other than antonyms. The antonyms appear to represent a special problem.

Method

The same test forms were used as in Experiment VII. The study list consisted of 50 single words (plus two filler words at the beginning and end of the series), 10 in each class, representing 10 of the pairs. These were in fact the same words as tested in Experiment VII. By using two different presentation lists, each of 20 words in each class served both as new words and as old words.

The *S*s were fully instructed concerning the study list and the method of testing. There were 24 completed records for each of the two forms for a total

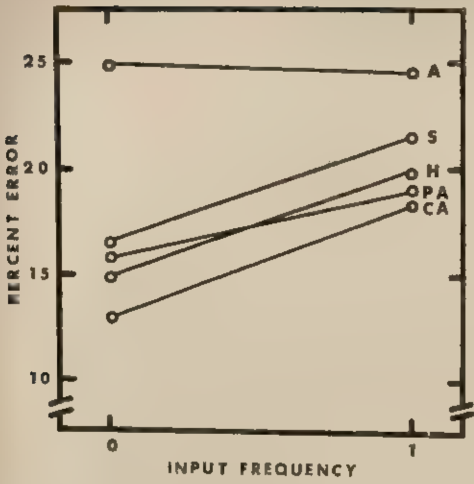


FIGURE 8. Recognition errors on single words following presentation of single words from five classes of pairs.

of 48. All procedures were exactly the same as in the previous studies.

Results

The essential results are shown in Figure 8, which is plotted in the same manner as the other recent graphs. It is to be noted that the results for the single words from the homonym pairs fall well within the limits set by the parallel associates, conceptual associates, and synonyms. This is to conclude, therefore, that there seems to be nothing peculiar about the words in the homonym pairs when recognition performance is observed for them as single words.

Figure 8 shows that the single words from antonym pairs produced many false alarms, as well as more misses than did the words from the other classes. Performance on the antonyms was largely responsible for the reliable class effect, F (4, 188) = 8.46, $p < .01$. There is a suggestion that a criterion difference might exist for the antonyms versus the other classes, but the interaction between old–new and type was less than one. It should be clear that the new items in the test were not, for any of the classes, the associated word as shown in Table 2. The false alarms cannot, therefore, be attributed to implicit associative responses elicited during the study trial which were represented in the new words on the test. If the false alarms are to be attributed to such implicit responses they arise from sources which cannot be easily identified. That some of the false alarms could be attributed to this mechanism is not to be denied, but it is not apparent why these should fall more heavily on the antonyms than on the other classes. The antonym problem will not be pursued further in this paper.

General Discussion

This research was initiated to determine the role which an association between a pair of words played in recognition. Given the knowledge of this role (if such existed) it would be possible to assess the degree to which frequency theory fails to accomodate it. The major purpose of this discussion is to make this assessment. The basic evidence from the experiments will be summarized initially.

1. Experiments I and II showed that parallel procedures in which recognition decisions were required in one case and frequency judgments in the other produced the same outcomes quantitatively. This was true for both absolute and relative judgments. Errors tended to fall on the same pairs for both types of judgments, and recognition errors were predictable from frequency judgments by combining the means and variance of the frequency judgments.

2. The paired-associate study showed that pairs on which few recognition errors were made were more easily associated than were pairs on which many errors were made. This would seem to implicate associability as a factor in recognition performance over and above frequency. Experiments III and IV supported this conclusion in showing that a pair of words, both shown on the study list but in different pairs, were less likely to be recognized than were intact pairs.

3. Experiment VI, however, in which pairs of culturally associated words were used, provided no consistent evidence that recognition performance was better for associated pairs than for nonassociated pairs. If one assumes that the association in culturally associated words is not qualitatively different from the associations established in the laboratory between initially neutral words, the findings with the culturally associated words are in direct conflict with those summarized in point 2 above.

4. Quite unexpectedly, criterion differences were found between frequency judgments and recognition decisions, the criterion being more lenient for the former. Recognition performance on homonym pairs also reflected a much more lenient criterion than was found for control pairs and culturally associated pairs.

The results of Experiments I and II make it difficult to turn abruptly away from frequency as the basic discriminative attribute for pair recognition. Therefore, the attempt has been made to develop a coherent account of the basic findings, using an expanded theory of frequency discrimination. This expansion is post hoc, and consequently must be viewed with skepticism, although the formulation as a whole does rest on frequency theory as developed in recent years. Before describing the account, it seems necessary to show how associative information or associative strength between words in a pair cannot be taken seriously as a direct explanatory mechanism for pair recognition.

Experiment VI showed that associated pairs and nonassociated pairs produced equivalent recognition performance. It would be quite possible to obtain such a result even if the recognition performance in the two cases was mediated by quite different types of information in memory. To be more specific, for the associated pairs the decisions may have been based exclusively on the associative relationship while the decisions for the nonassociated pairs may have been based on frequency information. Two considerations argue against the idea that the associative information is critical in the recognition of the culturally associated pairs of words. First, the new pairs for the associated pairs were also highly associated pairs. Therefore, some information other than associative information would be required to distinguish between the old and new pairs. Otherwise, the *S*s should have accepted any associated pair as having been included in the study list. Still, it might be argued (and this leads to the second point) that the single presentation of the associated pairs so increased the momentary situational strength of the association for the associated pairs that they were distinguishable on this basis from the control pairs even though the latter were strongly associated by cultural usage. This does not seem to be a likely possibility. In Experiment II, 12 pairs made up of unrelated words were presented twice during the study trial, and 12 pairs were presented three times. On the *yes–no* recognition test there were 5.3% misses for the former pairs, 3.5% for the latter. Although comparing error rates from experiment to experiment is risky because of many possible differences in subject samples, number of words presented and tested, and so on, these values must be considered appreciably less than the misses for the culturally associated pairs presented once. It does not seem appropriate to conclude that a neutral pair of words, presented twice, has a stronger associative link between them than does a pair of culturally associated words, such as *table–chair*, presented once. Based on associative strength alone, there is no reason to believe that the *S*s should have made any errors on the culturally associated pairs of words. Associative strength, at least as the term is understood as a consequence of multiple trials in the laboratory, does not seem to be directly involved at all in the recognition decisions. This conclusion was important in directing the explanatory efforts back toward frequency as the fundamental discriminative attribute for pair recognition.

In a recent study (Underwood & Zimmerman, in press), the evidence pointed toward the fact that each syllable of a two-syllable word gained a small amount of subjective frequency which was independent of the frequency information for the word per se. This is to say that the subunits of larger units may have frequency representation in memory. This abstractive nature of the assimilation of event frequency has long been known to be a characteristic of memory in the developmental history of the individual (Underwood, 1971). It now seems evident that this same feature obtains to some degree in the laboratory. The implication of this fact for frequency theory as applied to recognition of pairs is that the frequency information in memory will consist of information about each word, and also about the two words as a unit.

Given that frequency assimilation for a pair may involve three event frequencies, it is reasonable to ask what the optimal frequency should be for each event if recognition errors are to be minimized following a single presentation of a pair. Obviously, this arrangement would be one in which each unit has a subjective frequency of one, and the pair as a unit has a subjective frequency of one. All three sources of information would lead the *S* to respond with *yes*. A reduction in the subjective frequency for any one of the three events should reduce the likelihood of a correct

decision. Assumptions will now be made to reflect changes in the subjective frequency of each of the three events, hence changes in the likelihood of a correct decision.

It will be assumed that the stronger the association between two words *prior* to the study trial, the greater the likelihood that the subjective frequency of the pair as a unit will approach one, and the less the likelihood that the subject frequency of each word will approach one. The converse may be stated. If there is no associative relationship existing between two words prior to the study trial, the greater is the likelihood that the subjective frequency for the individual words will approach one, and the less the likelihood that the subjective frequency of the pair as a unit will approach one. In the extreme case, if no associative link is developed between the words in a pair during the study trial, the subjective frequency of the pair as a unit would be zero. The application of these notions to the data may now be examined.

1. In Experiment III, re-pairing old words on the test resulted in an increase in the errors. This would be accounted for on the grounds that a re-paired item has no subjective frequency for the pair as a pair. The decision must be made on the basis of the subjective frequency for the individual words. In Experiment IV, errors on the recognition of a single word from a pair presented for study increased (as compared with the presence of the intact pair), and the amount of increase was independent of the newness or oldness of the context word. The lack of pair-frequency information is believed responsible. In a further experiment, not reported in this series, individual words were presented for study. On the test, pairs of individual words were presented just as in Experiment IV and with *S* always making a recognition decision on the second word. Just as found by Thomson (1972), the characteristics (old or new) of the first word in the pair had no effect on the recognition of the second. Since there was no pair-frequency information induced during study, no loss of accuracy should have been found by prefixing either old or new words to the test word.

2. Pairs on which few recognition errors were made were more easily associated (in a paired-associate test) than were pairs on which many errors were made. The easy pairs were also judged to have higher frequency than were the difficult pairs. The easy pairs represent pairs which approximate as cl[illegible] as it is probably possible the maxin[illegible] frequency case. The pair was not asso[illegible]d prior to the study trial. But, for whate[illegible] reason, an association could be establishe[illegible]dily. In this process, each word will r[illegible]e a near maximum frequency (one), an[illegible] addition, the pair will approach this max[illegible]n because of the rapid formation of the as[illegible]ion.

3. Why were not cultural[illegible]ssociated words better recognized than [illegible]ssociated words? It is now presumed th[illegible]is result is a coincidence of the distributi[illegible] subjective frequency information. For [illegible]ssociated pairs, the decisions are primarily [illegible]d on pair frequency, for the nonassocia[illegible]pairs, on element or single-word frequency [illegible]It would be suspected that the small di[illegible]ences in errors among the four classes [illegible]sociates resulted from differences in the m[illegible]nitude of the subjective frequency for the s[illegible]le words.

It should be repeated that the f[illegible]lation is post hoc. However, it does have [illegible]umber of testable implications. Although [illegible]ese will not be examined here, it can be se[illegible] that the basic idea would be to evolve [illegible]cedures which would change the subjective [illegible]equency of one or more of the sources assum[illegible]d by the theory to be involved in the [illegible]ognition decisions.

The theory does not seem capable of handling cleanly the false-alarm data for N–O and O–N pairs in Experiments III and IV. All that can be said is that since subjective frequency should be present for one of the three events held important by the theory, the *S* is led to a *yes* response. This is not considered a satisfactory account for such false alarms.

Finally, the result for the homonym pairs provided a special case for frequency theory, a case that touches on several issues. Frequency theory makes no assumption concerning the awareness of the *S* that he is using frequency information in recognition decisions. In response to questioning, *S*s will sometimes respond that the correct item looked familiar, but more frequently given is the response: "It looked right." It will be remembered that the only distinctive difference found between recognition decisions and frequency judgments in Experiments I and II was that the *S*s set a more lenient criterion in making frequency judgments than in making recognition decisions. Now, when the *S* is presented homonym pairs he may, with some level of intention, make his decisions on the basis of the frequency

of the acoustic–articulatory responses he produced to a pair during the study trial. If this is the case, then whatever leads the S to set a lenient criterion for frequency judgments may likewise lead him to do the same in recognition of pairs of homonyms. That frequency of the acoustic–articulatory response is implicated seems clear by the results of Experiment VII. This study showed that relatively few misses were made on a single word from a homonym pair, a result that was not due to a lenient criterion. The S, in effect, had a subjective frequency of two for the single acoustic–articulatory response; for the other classes of items, the frequency would be appreciably less.

This series of experiments was initiated with the expectation that associative information in memory would cause a serious breakdown in the usefulness of frequency information as an explanatory mechanism. The evidence as viewed has led to the conclusion that it would be premature to abandon frequency theory in attempts to account for recognition performance for associatively related verbal units.

REFERENCES

Broder, P. Multiple-choice recognition as a function of the relative frequencies of correct and incorrect alternatives. Unpublished doctoral dissertation, Northwestern University, 1973.

Cofer, C. N., Segal, E., Stein, J., & Walker, H. Studies on free recall of nouns following presentation under adjectival modification. *Journal of Experimental Psychology*, 1969, **79**, 254–264.

DaPolito, F., Barker, D., & Wiant, J. The effects of contextual changes on component recognition. *American Journal of Psychology*, 1972, **85**, 431–440.

Ellis, H. C., & Shumate, E. C. Encoding effects of response belongingness and stimulus meaningfulness on recognition memory of trigram stimuli. *Journal of Experimental Psychology*, 1973, **98**, 70–78.

Light, L. L. Homonyms and synonyms as retrieval cues. *Journal of Experimental Psychology*, 1972, **96**, 255–262.

Light, L. L., & Carter-Sobell, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 1–11.

Thomson, D. M. Context effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 497–511.

Tulving, E., & Thomson, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–124.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

Underwood, B. J. Recognition memory. In H. H. Kendler & J. T. Spence (Eds.), *Essays in neobehaviorism.* New York: Appleton-Century-Crofts, 1971.

Underwood, B. J. Word recognition memory and frequency information. *Journal of Experimental Psychology*, 1972, **94**, 276–283.

Underwood, B. J., & Freund, J. S. Testing effects in the recognition of words. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 117–125.

Underwood, B. J., & Zimmerman, J. The syllable as a source of error in multisyllable word recognition. *Journal of Verbal Learning and Verbal Behavior*, in press.

Underwood, B. J., Zimmerman, J., & Freund, J. S. Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology*, 1971, **87**, 149–162.

Wolford, G. Function of distinct associations for paired-associate performance. *Psychological Review*, 1971, **78**, 303–313.

Wood, G. Associations and recognition learning. *Psychonomic Science*, 1969, **17**, 222–223.

(Received September 15, 1973)

Manuscripts Accepted for Publication in the

Journal of Experimental Psychology

Negative Recency in Initial Free Recall. Charles P. Thompson (Department of Psychol… Kansas State University, Anderson Hall, Manhattan, Kansas 66506), John M. Gardiner, and Am… Maskarinec.

Immediate and Delayed Outcomes: Learning and Recall of Responses. Alexander M. B… ld (Psychological Clinic, Indiana University, Psychology Building, Bloomington, Indiana 4… and Robert B. Meagher.

Odor Intensity and Pleasantness of Butanol. Howard R. Moskowitz (Pioneering Re… Laboratory, U. S. Army Natick Laboratories, Natick, Massachusetts 01760), Andrew Drav… nd Clifford Gerbers.

Comparison of Performance with Headphone and Free-Field Noise. L. R. Hartley … l Research Council, Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, England, … Carpenter.

Judgments of Pleasingness and Interestingness as Functions of Visual Complexity. P. P. … en (Department of Psychology, University of Keele, Keele, Staffordshire ST5 5BG, England).

Adaptation to Sensorimotor Conflict Produced by the Visual Direction of the Hand S… from the Cyclopean Eye. Hiroshi Ono (Department of Psychology, York University, 4700 Ke… t, Downsview 463, Ontario, Canada), and Robert G. Angus.

Organizational Attributes in List Acquisition and Retention. James W. Pellegrino (Learn… search and Development Center, University of Pittsburgh, University Drive, Pittsburgh, Pen… a. 15213).

Differential Recall as a Function of Socially Induced Arousal and Retention Interval. Ke… A. Deffenbacher (Department of Psychology, University of Nebraska at Omaha, P. O. Box 688, O… , Nebraska 68101), Gary J. Platt, and Mark A. Williams.

Effects of Irrelevant Color Changes on Speed of Visual Recognition following Short Rete… Intervals. Neal E. A. Kroll (Department of Psychology, University of California, Davis, C… nia 95616), M. H. Kellicutt, Raymond W. Berrian, and Alan F. Kreisler.

An Analysis of Proactive Inhibition in a Cued Recall Task. Linda R. Warren (Departmen… Psychology, Williams College, Williamstown, Massachusetts 01267).

Effect of Picture Word Transfer on Accuracy and Latency of Recognition Memory … C. Daniel (Department of Psychology, University of Arizona, Tucson, Arizona 85721), and Lo… Arthur.

Intrusion of a Thematic Idea in Retention of Prose. D. James Dooling (Department of Ps… logy, Kent State University, Kent, Ohio 44242) and Rebecca A. Sulin.

Articulatory and Constituent Phrases as Facilitators of Word Identification Decisions. …elen Altman Klein (Department of Psychology, Oakland University, Rochester, Michigan 48063), Gary A. Klein, and Donald C. Hildum.

Order Information in Short-Term Memory. Douglas K. Detterman (Department of P… logy, Case Western Reserve University, Cleveland, Ohio 44106) and Jane Brown.

Heart Rate and Muscle Tension Correlates of Conditioned Suppression in Humans. M. G… (Department of Psychology, University of Newcastle, Shortland, New South Wales 2308, Austr… , Janice A. Di Giusto, and Eros L. Di Giusto.

Effects of Formal Similarity: Phonetic, Graphic, or Both? Douglas L. Nelson (Department of Psychology, University of South Florida, Tampa, Florida 33620), David H. Brooks, and Richard C. Borden.

The Image as Mediator in One-Trial Paired-Associate Learning: III. Sequential Functions in Serial Lists. B. R. Bugelski (Department of Psychology, State University of New York at Buffalo, 4230 Ridge Lea Road, Buffalo, New York 14226).

Role of Rehearsal in Long-Term Retention. Daniel L. Roenker (Department of Psychology, Iowa State University, Old Botany Hall, Ames, Iowa 50010).

Proportion of Unitization as an Index of Cued Recall Level. Michael A. Karchmer (University of Kansas Medical Center, Kansas Center for Mental Retardation and Human Development, Kansas City, Kansas 66103).

Testing Tapping Time-Sharing. Barry H. Kantowitz (Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907) and James L. Knight, Jr.

Mechanism of Input Selection in the Selective Perceptual Processing of the Accepted Message in a Dichotic Auditory Presentation. Richard J. Rindner (34 Metropolitan Oval, Bronx, New York 10462).

Selective Backward Masking with an Unpredictable Mask. Philip M. Merikle (Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada).

Facilitation of Sequential Short-Term Memory with Pictorial Stimuli. Judith P. Allik (Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260) and Alexander W. Siegel.

Memory for Random Shapes: A Dual-Task Analysis. David W. Martin (Department of Psychology, New Mexico State University, Box 3452, Las Cruces, New Mexico 88003) and Richard T. Kelly.

Visual Masking by Light Offset: An Experiment in Reply to Hogben and Di Lollo. R. James Holzworth (Psychology Department, Bowling Green State University, Bowling Green, Ohio 43403) and Michael E. Doherty.

Influence of the Shape of the Receptor Organ on the Horizontal Vertical Illusion in Passive Touch. T. S. Wong (Department of Psychology, University of Waikato, Hamilton, New Zealand), R. Ho, and J. Ho.

Effects of Stimulus Complexity on the Perception of Brief Temporal Intervals. H. R. Schiffman (Department of Psychology, Rutgers—The State University, New Brunswick, New Jersey 08903) and Douglas J. Bobko.

Relative Effects on Performance and Motivation of Self-Monitoring Correct and Incorrect Responses. Terry C. Wade (Department of Psychology, University of Utah, Salt Lake City, Utah 84112).

CONTENTS—(*Continued*)

9/8 74

Journal of Experimental Psychology

CONTENTS

ARTICLES

(Continued on inside back cover)

American Psychological Association

Vol. 102, No. 6 June 1974

Beginning in 1975 the *Journal of Experimental Psychology* will appear as four [illegible] edited and distributed sections.

The Journal of Experimental Psychology: Human Learning and Memory (Lyle E. Bourne, Jr., Editor) will publish experimental studies on fundamental acquisition, retention, and transfer [illegible] in human behavior.

The Journal of Experimental Psychology: Human Perception and Performance (Michael I. Posner, Editor) will publish experimental studies designed to foster understanding of [illegible] operations and their relation to experience and performance.

The Journal of Experimental Psychology: Animal Behavior Processes (Allan R. Wagner, Editor) will publish experimental studies concerning the basic [illegible] perception, learning, [illegible], and performance, especially as revealed in the behavior of [illegible] animals. The studies should make significant contributions to general behavior theory.

The Journal of Experimental Psychology: General (Gregory A. Kimble, Editor) will publish experimental articles in any of the [illegible] represented by the above [illegible] when the articles involve a longer, more integrative report leading to an advance in knowledge that is [illegible] of [illegible] community of experimental psychologists. The section will include but not be limited to articles such as those that have appeared as *Journal of Experimental Psychology* monographs and as chapters [illegible] books of "advances." Republication of a limited number of data may be permitted if necessary to [illegible] the article complete and definitive.

In some instances it may be difficult for the author to decide whether one of the above [illegible] sections or the *Journal of Comparative and Physiological Psychology* is most appropriate for a given [illegible]. The author should weigh the essential implications of the research rather than the specific [illegible] subject population employed. Is the central message concerned with the issues and concepts of [illegible] theory (*JExP*) or with the biological substrates of behavior (*Journal of Comparative and Physiological Psychology*)? Is the study primarily addressed to theory of human behavior (*JExP: Human Learning and Memory* or *JExP: Human Perception and Performance*) or to theories equally or more relevant to infrahuman subjects (*JExP: Animal Behavior Processes*)? Is the major emphasis on the acquisition, retention, or transfer of new information, concepts, or problem-solving strategy (*JExP: Human Learning and Memory*) or on the organization or control of structures or mental operations, where the acquisition of new skills is only incidental (*JExP: Human Perception and Performance*)?

Manuscripts should be sent directly to the Editor of the relevant section at the address listed below. In cases where an Editor believes a manuscript is clearly more appropriate for an alternative journal of the American Psychological Association, he may redirect the manuscript with the approval of the author. A final action taken by one of the Editors will be final for all sections of *JExP*.

Manuscripts should be submitted in triplicate and according to the instructions contained in the *Publication Manual* of the American Psychological Association.

Authors should note that limitations on total pages make it necessary to restrict publication to those papers that in the opinion of the Editors make a notable contribution to the literature. Preference can be expected for sets of interrelated experiments that are definitive or provocative.

Addresses for the 1975 Editors-Elect are as follows. *JExP: Human Learning and Memory*, Lyle E. Bourne, Jr., Department of Psychology, University of Colorado, Boulder, Colorado 80302; *JExP: Human Perception and Performance*, Michael I. Posner, before June 1, 1974, same as Allan R. Wagner, after June 1, Department of Psychology, University of Oregon, Eugene, Oregon 97403; *JExP: Animal Behavior Processes*, Allan R. Wagner, Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510; *JExP: General*, Gregory A. Kimble, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1974: $60 domestic, $61 foreign, $6 single issue. APA member rate: $25. Write to Subscription Section, APA.

Reprints: Authors may order reprints of their articles from the printer when they receive galley proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P. O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. Undelivered copies resulting from address changes will not be replaced; second-class mail is not forwarded unless subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Published monthly in two volumes, six issues per volume, by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Washington, D.C., and additional mailing offices.

Journal of Experimental Psychology

Volume 102, 1974

PUBLISHED BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 SEVENTEENTH ST., N.W., WASHINGTON, D. C. 20036

CONTENTS OF VOLUME 102

Journal of

Experimental Psychology

Vol. 102, No. 6 JUNE 1974

SIMULTANEOUS DISCRIMINATIONS: TWO SEPARATE TYPES OF CONTROL [1]

MICHAEL D. ZEILER [2]

Emory University

Children and retarded adults were given four different learning conditions and a series of transfer tests to determine which stimulus aspect controlled performance in a simultaneous discrimination. An adequate description of the discrimination performance entailed the recognition that different individuals displayed equal accuracy but were not controlled by the same stimulus aspect. Eight of the *S*s responded on the basis of stimulus relationships, but the other 14 *S*s were not sensitive to those aspects of the stimuli. One *S* was controlled by irrelevant stimulus properties; the others demonstrated that choices depended on the size differences among the stimuli of each three-member set and the differences between the learning and the transfer sets. A model proposed to explain these performances could describe the probability of responding in each individual. Thus, individuals showed two types of control, one characterized by sensitivity to stimulus relations and the other described by the model.

Two problems confronted by theoretical accounts of discrimination learning are specifying the controlling aspect of stimuli and describing the way the discriminations are learned. Stimuli can differ along a number of dimensions, e.g., two visual stimuli can have different colors, forms, sizes, spatial locations, etc. An individual whose responding varied depending on which of the two stimuli was presented might be under the control of any or all of these aspects, but which one was relevant could not be determined from the sheer occurrence of differential responding. The achievement of differential responding also does not indicate the way the problem was learned (e.g., incrementally, all-or-none learning, etc.). In fact, it has been demonstrated that similar criterion performances can be attained in a variety of ways in the same task (Gholson, Levine, & Phillips, 1972), thereby indicating that no single-process model can provide a comprehensive account. The same conclusion can be reached about models attempting to explain which aspect of the stimuli controls performance in even a simple situation involving stimuli differing on only a single dimension.

One reason that research on controlling stimulus aspects has been insensitive to the possibility of multiple sources of control has been that data were averaged over groups of *S*s (cf. Reese, 1968). Theories were devised to explain these averaged

[1] This research was supported by Research Grant HD-05752 from the U.S. Public Health Service. The author would like to thank Howard Rollins for his many helpful comments and criticisms.

[2] Requests for reprints should be sent to Michael D. Zeiler, Department of Psychology, Emory University, Atlanta, Georgia 30322.

data on the implicit assumption that the means estimated the performance of individual *Ss* (e.g., Zeiler, 1963). However, the detailed evaluation of individuals revealed that the means were not an adequate approximation and consequently the theories were not explaining individual performance corrected for sources of uncontrolled error by averaging across *Ss*. For example, some *Ss* were always controlled by the relational characteristics of two simultaneously presented stimuli, whereas others were sensitive to absolute stimulus sizes rather than to relationships (Kuenne, 1946; Porter, 1969). With situations involving three stimuli with the middle-size member of an original training condition as the correct set member, some *Ss* were always controlled by the middle-size relation, but others responded to the middle-size stimulus in transfer tests only if the new stimuli were very similar to the originals (Zeiler & Friedrichs, 1969; Zeiler & Salten, 1967). In fact, the same individuals have even revealed both types of control simultaneously (Zeiler, 1964).

The existence of two controlling stimulus aspects means that the averaging of the performances of different individuals must yield misleading data. A more profitable approach might be to recognize the heterogeneity and attempt to formulate a theoretical description of each type. Subsequent steps would be to try to understand what variables are responsible for determining which aspect will predominate, as well as to analyze how both are combined in the behavior of adult humans. The present experiment was intended to provide the first step, i.e., the description of the two types of control emerging with three stimuli varying in size.

One type of control seems straightforward. Some *Ss* are controlled by the relationships of one stimulus to another. However, the second form of control cannot be analyzed so simply. In this case, individuals sometimes choose the stimulus having a certain relationship, but at other times do not. This sort of profile is predicted by the ratio theory (Zeiler, 1963). According to the ratio theory model, each stimulus is defined by the ratio of [illegible] the prevailing stimulus contex[illegible]daptation level). The training adaptation level is determined by the training stimuli alone (the geometric mean of their areas), and the test adaptation level is the weighted sum of the geometric mean of the test stimuli and the geometric mean of the training stimuli. In the t[illegible] will prefer the set member having the [illegible] closest to that of the positive stimulus (S+) in training.

With three stimuli differing in size by equal units and the middle-sized stimulus as S+ during the training phase, which test stimulus will be preferred depends on the similarity of the stimuli within each set and the similarity of the training stimuli to the test stimuli. As the difference *between* the training and the testing conditions increases from zero, the test adaptation level changes in such a way that the model predicts that the preferred stimulus should shift from the middle-sized test set member to a different stimulus. The stimulus differences *within* the test sets are also critical. If a particular stimulus is the middle-sized member of a test set and is preferred, increases in the differences between it and the accompanying larger and smaller set members will not affect the preferences. However, if either the large or the small stimulus is preferred, increasing within-set differences will eventuate in a shift to preference of the middle-sized member. Thus, the model predicts that between- and within-set differences operate in conjunction to determine choices in the tests.

None of the research up to this point that has studied the performances of individual *Ss* can evaluate the accuracy of this prediction, because no experiment has manipulated within- and between-set stimulus differences within the same individuals. The purpose of the present experiment was to do this. Because normal adults have been shown to switch the basis of their choices as a function of between-set similarity (Zeiler, 1964), young normal children and retarded adults were used as *Ss* in order to assure the observation of independent patterns of transfer performance.

These Ss were expected to reveal one or the other of the two components that interact to determine the adult behavior.

Method

Subjects. The Ss were 16 5- and 6-yr. old children (8 boys and 8 girls) attending a summer day camp for normal children and 6 institutionalized retarded adult women ranging in age from 16–30 yr. There were no current intelligence test scores available for any of the Ss, but a variety of tests conducted on admission reported that the IQs for the retarded women ranged 40–68.

Apparatus. The stimuli were two identical sets of 3-mm.-wide black lines, one set used for training and the other for the transfer tests. The lines are designated in order of length; the smallest (1) was 18-mm. long, and each successive line increased in length by a factor of 1.2:1. Each line was cut from glossy drafting tape and was mounted on a 5 × 25 cm. block of 6-mm.-thick white Masonite. A piece of 6-mm.-thick clear plastic (Plexiglas) was cut to the same size as the Masonite block and glued over it. The lower edge of each line was 1 cm. from the bottom of the block.

Stimuli were displayed on a turntable that could be rotated to interpose a screen between S and the blocks. The three blocks used on each trial were arranged so that the lines were vertical with respect to S and bottoms of the lines were aligned. The long edges of the blocks touched, and the bottoms faced S.

The stimuli comprising the training and test sets are shown in Table 1. The combinations of sets differed from each other in the similarity of set members to each other (within-set difference), and the four test sets varied in their similarity to the training set (between-set difference). The same stimulus (Line 5, length = 36 mm.) was the middle-sized and positive member of the four different training sets.

Procedure. A noncorrection procedure was used throughout. On each trial three blocks were exposed until S made a choice. Then the screen was interposed between S and the blocks, and E made any necessary changes in position and/or blocks. Stimulus positions varied according to a predetermined irregular sequence.

In Phase 1 the child found a plastic chip under the middle-sized stimulus of the training set. The three blocks were in view when S came into the experimental room. Instructions were as follows.

> See this chip? See these blocks? I am going to hide this chip under a block. You find it. Try to find as many as you can. Choose only one block on each turn.

Phase 1 ended when S made five successive choices of the middle-sized block.

In Phase 2, the chips were discontinued for the normal children (they were continued for the retarded Ss), but a penny was given following every fifth correct response. The S was told that a penny would be hidden instead of the chip and could be earned by choosing the correct block a few times in a row. The S was also told:

> Only one block is right. The same block is always right. I always hide the penny under the same block.

Phase 2 concluded with 20 successive correct choices of the middle-sized stimulus.

Phase 3 was the testing phase. Each of the four test sets associated with each training set was presented for four trials. The order of test set presentation was random with the limitation that each set appear in each block of four test trials. A test trial appeared after six or seven successive correct responses with the training set. Pennies continued to be given for every fifth correct response with the training set, but choices had no consequences on test trials.

After the three phases were completed with the first training set and its associated test sets, a different training set was introduced and the procedure was repeated. This sequence of conditions was continued until each S had been given the four different training sets and the four appropriate test sets. Thus, after the first four conditions, S had been trained with the four training sets and had been given four trials with each of the test sets. There were then four more conditions which repeated the first four, thereby yielding a total of eight trials with each test set. The order of conditions was randomized with the limitation that each of the four training sets be used once before any was used twice—each S had a different order. Within each condition, the order of the test sets was also randomized with the same limitation.

In the first condition Phases 1 and 2 were conducted in a single session, and all of Phase 3 was conducted in a single session. For the retarded Ss Phases 1 and 2 sometimes required as many as 3 sessions to attain criterion performance in each. In the seven subsequent conditions, Phase 1 was omitted, and Phases 2 and 3 were conducted in a single session. The experiment, therefore, involved 9 sessions for the normal children and between 9 and 11 sessions for the retarded adults.

A display of toys and candy was located on shelves in the experimental room. Also, there was a box

TABLE 1

Stimulus Sets

Within-set difference	Training set	Training–test difference			
		1.2 :1	1.4 :1	1.7 :1	2.1 :1
1.2:1	4-5-6	5-6-7	6-7-8	7-8-9	8-9-10
1.4:1	3-5-7	4-6-8	5-7-9	6-8-10	7-9-11
1.7:1	2-5-8	3-6-9	4-7-10	5-8-11	6-9-12
2.1:1	1-5-9	2-6-10	2-7-11	4-8-12	5-9-13

containing trinkets located on the table holding the turntable. The trinkets were valued at 1 cent (normal *Ss*) or 1 chip (retarded *Ss*) each, whereas the toys and candy ranged from 5 to 20 cents or chips. On any given day, *S* could buy toys and/or candy if he or she had sufficient money or chips, buy trinkets, keep the pennies or chips, and/or save as many as desired to purchase more expensive toys.

Results and Discussion

Performance with training sets. The mean number of errors made in attaining criterion in Phase 1 with the first training condition was 15.6 for the normal children and 38.3 for the retarded adults. This difference was significant ($p < .01$, Mann–Whitney U test). No errors were made in Phase 2 or in the training trials of Phase 3. Errors in Phase 1 did not differ significantly ($p > .10$, Kruskal–Wallis test) depending on the particular training set either for all of the *Ss* combined or for the normal children aad retarded adults separately. The means and ranges of errors were as follows: Set 4-5-6, $M = 25.3$, range = 3–65; Set 3-5-7, $M = 17.0$, range = 2–36; Set 2-5-8, $M = 20.0$, range = 2–45; Set 1-5-9, $M = 26.3$, range = 6–50.

In training subsequent to the first condition, errors occurred only with Set 4-5-6, with a mean of 2.3 (range 0–12) for the normal children and a mean of 13.3 (range 3–27) for the retarded adults. This difference did not attain statistical significance ($p > .05$, Mann Whitney U test). No errors occurred with any other training set. As in the first condition, there were never any errors with the training trials of Phase 3. Interpolated test trials in which responses were not reinforced had no apparent effect on the accuracy of the training performance.

Performance with test sets. The *Ss* had 8 trials with each test set under each of the four training conditions. Eight normal children (five girls and three boys) chose the middle-sized test set member in all 32 test trials. Consequently, their behavior fits the hypothesis that they were controlled by the middle-size relation. None of the remaining *Ss* demonstrated equivalent performance.

One normal child [illegible] different choice on each exposure to [illegible] iven test set and was able to choose S+ [illegible] n when the lines were covered. Verbal [illegible] rt and selective covering of various pa[illegible] he training blocks revealed that she [illegible] ponding to a discoloration on the block [illegible] er than to the line. Since this disc[illegible] n was not present on any test stim[illegible] lock, the responses during the tests [illegible] istributed at random.

The remaining seven norm[illegible] ldren and all retarded adults were sen[illegible] o changes in the sizes of the lines. [illegible] e was no relation between the perfor[illegible] e in original learning and the type of [illegible] patterns revealed by the normal c[illegible] Those children who did not reveal [illegible] choice of the middle-sized test stimu[illegible] d not be differentiated from those w[illegible] on the basis of the number of errors [illegible] riginal learning phases; they range[illegible] om the fastest to the slowest learner[illegible]

Figure 1 indicates the fr[illegible] cy of choices of the middle-sized t[illegible] imulus for each of these 13 *Ss*. All of [illegible] ponses not occurring to the middle-siz[illegible] timulus were to the smallest stimulus [illegible] he test set, i.e., there were no choices of [illegible] largest test set member. In every case, the frequency of choices of the middle-sized test set member declined and the frequency of choices of the smallest stimulus increased as the differences between the training and the test sets increased. The gradients of choice of the middle-sized stimuli typically were higher with each progressive increase in within-set stimulus difference, although there were occasional reversals.

Although there was only one instance of 2 *Ss* revealing the same choice gradients in all of the conditions (C_{16} and C_{20}), there were several in which 2 or more *Ss* revealed the same gradient in a given condition of within-set difference. In fact, 8 *Ss* revealed the identical gradient with the within-set difference of 1.2:1. The heterogeneity among the 13 *Ss* was larger with increased within-set stimulus differences.

Quantitative analysis. The dependence of choices on the combined effects of within- and between-set stimulus differences

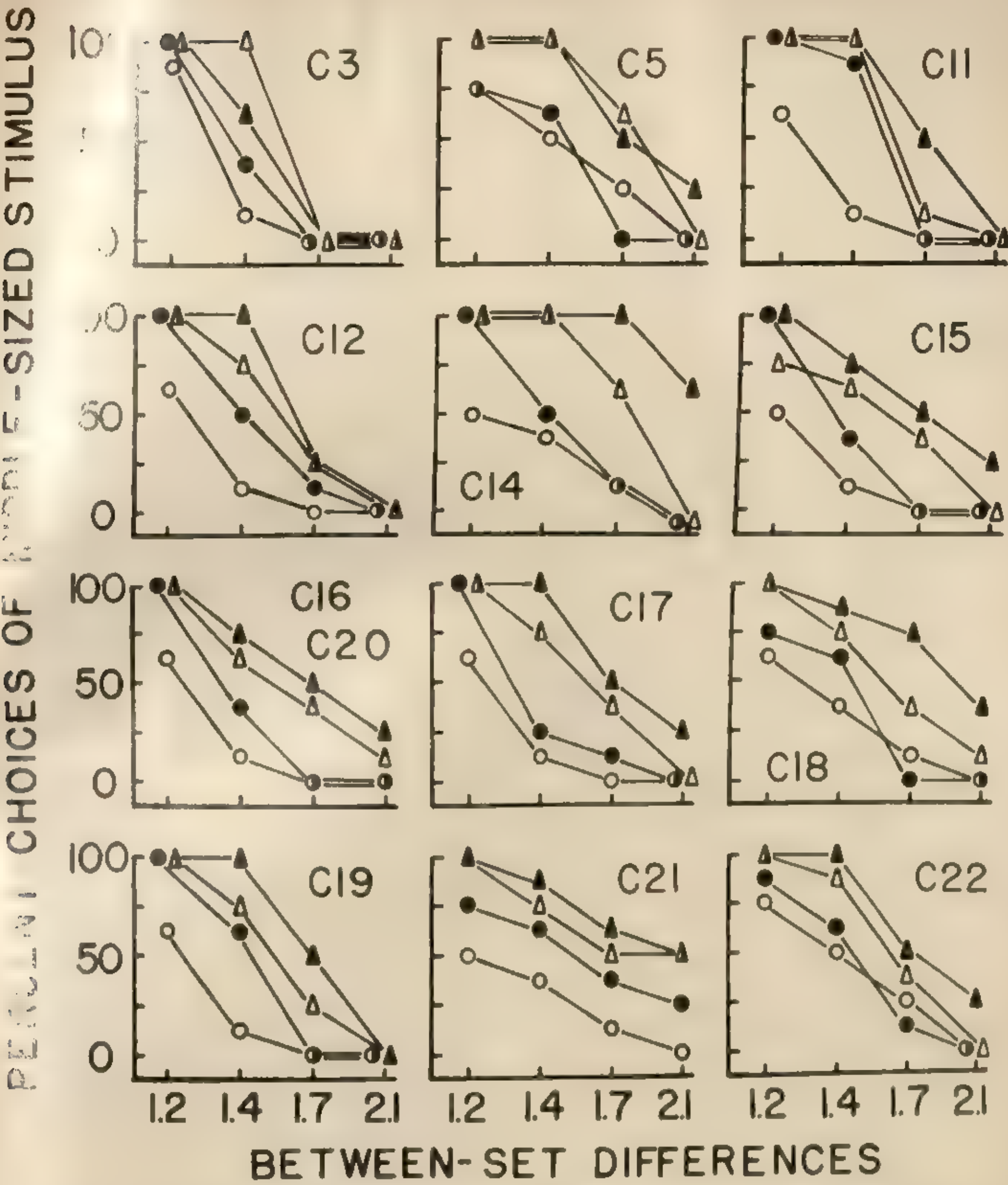


FIGURE 1. Percentage of choices of the middle-sized stimulus in each test set (The data are shown separately for each S. Open circles, within-set difference = 1.2:1, filled circles, within-set difference = 1.4:1; open triangles: within-set difference = 1.7:1; filled triangles: within-set difference = 2.1:1.)

was predicted by the ratio theory model (Zeiler, 1963). The two general predictions were confirmed: (*a*) at any one level of between-set difference, choices of the middle-sized stimulus increased and choices of the smallest set member decreased with increases in within-set difference; and (*b*) at any one level of within-set difference, choices of the middle-sized set member increased and choices of the smallest stimulus decreased with increases in between-set differences. It also appeared that a modified

TABLE 2

BEST FITTING CONSTANTS

Subject	Within-set difference							
	1.2:1		1.4:1		1.7:1		2.1:1	
	b	y	b	y	b	y	b	y
C_1	18.0	.60	10.0	.40	8.0	.30	12.0	.20
C_6	10.0	.70	10.0	.50	12.0	.40	6.0	.40
C_{11}	12.0	.60	12.0	.30	12.0	.40	8.0	.30
C_{12}	12.0	.60	8.0	.50	6.0	.40	6.0	.20
C_{14}	10.0	.50	8.0	.50	10.0	.50	6.0	.60
C_{15}	12.0	.60	10.0	.40	4.0	.30	4.0	.30
C_{16}	12.0	.60	10.0	.40	6.0	.40	4.0	.30
C_{17}	12.0	.60	10.0	.40	6.0	.40	6.0	.40
C_{18}	8.0	.60	10.0	.50	6.0	.50	12.0	.60
C_{19}	12.0	.60	10.0	.50	6.0	.40	8.0	.30
C_{20}	12.0	.60	10.0	.40	6.0	.40	4.0	.30
C_{21}	8.0	.60	4.0	.60	4.0	.50	4.0	.40
C_{22}	10.0	.70	8.0	.50	8.0	.40	6.0	.40
M	11.4	.61	9.2	.45	7.2	.41	6.6	.36

form of the model could describe the probabilities of response in the 13 *Ss* sensitive to within- and between-set differences. The revised model consists of seven assumptions. Let Z_{oi} = area of training stimulus i ($i = 1, \ldots, m$); Z_{oc} = area of positive stimulus in training; Z_{ti} = area of test stimulus i ($i = 1, \ldots, n$); P_{ti} = probability of response to Z_{ti}; and y and b = constants. The seven assumptions are as follows:

Training adaptation level (A_o):

$$\log A_o = \frac{1}{m} \left(\sum_{i=1}^{m} \log Z_{oi}\right). \quad [1]$$

Test adaptation level (A_t):

$$\log A_t = y\left(\frac{1}{n}\left(\sum_{i=1}^{n} \log Z_{ti}\right)\right) + (1 - y) \log A_o. \quad [2]$$

Positive stimulus ratio (X_{oc}):

$$X_{oc} = \frac{Z_{oc}}{A_o}. \quad [3]$$

Test stimulus ratios (X_{ti}):

$$X_{ti} = \frac{Z_{ti}}{A_t}. \quad [4]$$

Probability of response:

$$D_{ti} = |X_{ti} - X_{oc}|, \quad [5]$$

$$V_{ti} = e^{-bd_{ti}}, \quad [6]$$

$$P_{ti} = \frac{V_{ti}}{\sum_{i=1}^{n} V_{ti}}. \quad [7]$$

Equations 6 and 7 were not contained in the original ratio theory: they represent two new assumptions and the addition of Luce's (1959) choice axiom to the model. The adoption of this axiom is not intended to imply that it is uniquely applicable to the present data (cf. Friedman, [illegible], & Padilla, 1968); it was chosen simply as a convenient technique for demonstrating the possibility of describing response probabilities within the context of the ratio theory.

For each *S* values of y and b were found that could describe the obtained probabilities of responding for the four test sets used under each level of within-set stimulus difference. Since $p = 0.0$ for the four largest stimuli of the test sets, these probabilities were excluded. The probabilities predicted by each pair of y and b values were compared with those obtained to determined which pair produced the smallest deviation. The search for the best was conducted by incrementing y in steps of .10 and b in steps of 2.0 until the average of the squared deviations for the four test sets was less than .01. The smallest values of the constants that produced deviations of this magnitude are shown in Table 2. (Note: once the b value that produced this small a deviation was found, y was set at the value that produced the least deviation.) Because there were 13 *Ss* × 16 Test Sets × 2 Stimuli in each, it is most awkward to try to present all of the expected probabilities. It is perhaps sufficient to indicate that no expected and obtained response frequency to a given stimulus deviated by as much as one response.

The constants varied among and within *Ss*. Since both tended to decrease as within-set differences increased, the constants seemed not to be arbitrary but rather to be a function of the particular stimuli

used. constants may also be a function dividual differences; however, their ricted range implies that such diff es may be less important than the stim us variables in determining the val s.

conclusion, these data indicate that the behavior of individuals in a three-mulus simultaneous discrimination can be classified into two general types. One involves control by stimulus relationships (although it may be worth noting that uni-rm choice of the middle-sized stimulus is predicted if y is 1.00); the other is described by the ratio theory model. Furthermore, it appears possible to provide a quantitative count of performance over a range of ining and testing conditions. Less clear this time is what variables are respon-ible for establishing one or the other type of control.

REFERENCES

Friedman, M. P., Rollins, H., & Padilla, G. The role of cue validity in stimulus compounding. *Journal of Mathematical Psychology*, 1968, **5**, 300–310.

Gholson, B., Levine, M., & Phillips, S. Hypotheses, strategies, and stereotypes in discrimination learning. *Journal of Experimental Child Psychology*, 1972, **13**, 423–446.

Kuenne, M. R. Experimental investigation of the relation of language to transposition behavior in your children. *Journal of Experimental Psychology*, 1946, **36**, 471–490.

Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

Porter, W. L. Absolute versus relational responding in children. *Journal of Experimental Child Psychology*, 1969, **7**, 485–491.

Reese, H. W. *The perception of stimulus relations: Discrimination learning and transposition*. New York: Academic Press, 1968.

Zeiler, M. D. The ratio theory of intermediate size discrimination. *Psychological Review*, 1963, **70**, 516–533.

Zeiler, M. D. Transposition in adults with simultaneous and successive stimulus presentation. *Journal of Experimental Psychology*, 1964, **60**, 103–107.

Zeiler, M. D., & Friedrichs, A. G. Absolute and relational discriminations involving three stimuli. *Journal of Experimental Psychology*, 1969, **82**, 448–452.

Zeiler, M. D., & Salten, C. S. Individual gradients of transposition and absolute choice. *Journal of Experimental Child Psychology*, 1967, **5**, 172–185.

(Received August 17, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 948–953

AN ANALYSIS OF INTERSENSORY TRANSFER O[illegible][1]

RONALD W. SHAFFER[2] AND HENRY C. ELLIS

University of New Mexico

The relative contributions of four sources of transfer in intersensory [illegible] between vision and touch were examined. Following a training task i[illegible] tifying visually presented polygons or histoforms, six groups learned t[illegible] a motor switching response to tactually presented polygons in a tr[illegible] task. Transfer performance showed reliable evidence for the transf[illegible] forward associations, minimal evidence for the transfer of stimulus [illegible] and no evidence for two types of nonspecific transfer.

Learning to make specific identifying responses to different stimuli in a training task is known to affect performance in a subsequent transfer task where the "same" stimuli are presented via different sensory modalities (e.g., Bjorkman, Garvill, & Molander, 1965; Gaydos, 1965; Holmgren, Armoult, & Manning, 1966; Sinha & Sinha, 1960). Such transfer has been termed intersensory transfer and several hypotheses have been advanced to account for its occurrence (e.g., Gibson, 1969, chap. 11; Wright, 1970). While intersensory transfer has been repeatedly demonstrated, very few studies have provided the kinds of operations that allow the partitioning of contributing sources of transfer. The present experiment provides controls not previously contained in intersensory transfer studies (e.g., Gaydos, 1956; Holmgren et al., 1966; Sinha & Sinha, 1960), thereby allowing isolation of sources of transfer specifically related to stimulus modality.

Two potential sources of specific transfer are of immediate interest in the intersensory transfer paradigm: the transfer of forward associations and of stimulus learning, i.e., learning how to discriminate among the stimuli. Transfer of forward associations with unfamiliar form stimuli can imply the operation of a stimulus encoding system superordinate to the separate sensory systems, since it would suggest that *S*s could "recognize" the same stimul[illegible]ed via a second sensory system. B[illegible] stimuli for which *S*s could have ha[illegible]revious opportunities to learn medi[illegible]ponses to their intersensory equiva[illegible]an be determined whether this kind [illegible]coding system need be advanced.

If, after learning to associat[illegible]tifying responses with stimuli, *S*s sh[illegible] performance decrement when tran[illegible] to a new modality, then the functio[illegible]ivalence of the intersensory stimu[illegible]d be evident. However, less than pe[illegible]ransfer performance requires controls [illegible]rtial out sources of transfer other tha[illegible]at of forward associations. The si[illegible]ntrol groups included in the present stud[illegible]ermit the separation of forward associations and the second source of specific transfer, stimulus learning, while additional controls allow the separation of two nonspecific sources of transfer.

The transfer of stimulus learning would suggest that the *changes* in the way stimuli are encoded are not specific to the particular sensory system under which they were learned. This finding would follow from the assumption that there are amodal sources of information (Gibson, 1969, chap. 11), and that learning to respond to these amodal stimulus attributes should modify how the same stimuli are encoded when presented to another sensory system.

In addition to converging on two specific sources of transfer, the present study provides the conditions for assessing nonspecific transfer in an intersensory transfer paradigm. It is common practice in studies of verbal transfer to use items for the nonspecific control that are independent of

[1] This research was supported by National Science Foundation Grants GB-7296 and GB-27413X to the second author.

[2] Requests for reprints should be sent to Ronald W. Shaffer, who is now at the Department of Psychology, Western Washington State College, Bellingham, Washington 98225.

specific list content, but are nevertheless sampled from the same class of stimuli as the other training conditions. For example, if adjectives are used as first-list items, then the baseline nonspecific control would also employ adjectives. While the dependency of the class of stimulus items on the amount of nonspecific transfer with verbal items has been demonstrated by Postman and Schwartz (1964), this variable has not been directly investigated with "nonverbal" form stimuli. It is not apparent, therefore, whether the degree of nonspecific transfer will remain invariant across different classes of nonverbal first-list stimuli. Moreover, the potential role of the *nature* of the form stimuli to be used for nonspecific controls becomes magnified when concern is focused on stimulus encoding changes relevant to the specific form stimuli used in the training–transfer task (i.e., stimulus learning), as opposed to *general* encoding changes particular to the class of form stimuli used but not to specific members within that class. Accordingly, several nonspecific controls are contrasted to the conditions given exactly the same forms in the visual training and tactual transfer task: Two groups receiving visual training forms from the same class (WC), and two groups receiving visual training forms from a completely different class (BC), which are of comparable ease with regard to associative learning.

Method

Stimuli and apparatus. Three sets of six forms were cut from 3.2-mm.-thick black rubber tile and mounted on white translucent plastic squares. Twelve of the stimuli were 12-point constrained metric polygons (CMPs) sampled from the population of shapes described by Thurmond and Alluisi (1967). The CMPs were constructed from a radiating matrix of 12 "lines" of points with a 30° angle between each successive radii of lines and with 12 points along each line. Sampling by constrained rules simply means that a particular point on any one line of points can appear once and only once in each form. The difference between successive points was arbitrarily set at 3.175 mm.

Another set of six stimuli (used for the BC nonspecific control) were six-column constrained metric histoforms (CMHs), also described by Thurmond and Alluisi (1967). The sides of the matrix from which the CMHs were generated were set at 7.62 cm. so that the CMH would approximate the overall size of the CMP. This particular subset of CMHs was chosen because pilot data had indicated that visual learning rates (training) would be comparable to learning rates with CMP stimuli.

The apparatus was constructed so that any particular form could be either viewed or tactually manipulated. The *S*s were seated in front of a 55.8 × 55.8 cm. plywood display panel painted flat black. Tactual manipulation was achieved by *S* reaching under a hood located at the bottom of the panel and tracing the perimeter of a particular form with his forefinger. The plastic square mounts were inserted individually by *E* into a holder located horizontally under the hood. The raised forms were used for visual presentation by inserting them behind a piece of 10.1 × 12.7 cm. one-way glass located directly above the hood. A rectangular tunnel was constructed directly behind the one-way glass. The pieces of plastic with the mounted forms were inserted into a tray slid through an opening in the tunnel by *E*. Hunter decade timers were used to control the duration of a 100-w. bulb located behind the slide opening. When the bulb was activated *S* was able to view the form through the one-way glass. Directly above the one-way glass was a window with a white paper insert which served as a rear-projection screen for the letter responses used in the training task.

The six letters (A, E, H, N, O, and R) used as the verbal response terms in the training task for three of the groups were projected on the rear of the paper via a Kodak Carousel 700 projector. Located directly above the verbal response "window" were six horizontally arranged spring-loaded toggle switches to which *S* responded in the training task for three groups and in the transfer task for all groups. Over each switch was a circular green confirmation light which, at the end of the anticipation interval, informed *S* as to the correct response. The *E*'s control panel contained switches to program the confirmation lights, a display panel to indicate which switch was chosen by *S*, and a master switch to initiate the presentation of each paired associate.

Experimental design. Ninety-eight undergraduate *S*s were assigned to each of seven treatment conditions. The *S*s were randomly assigned to conditions in blocks of 7, with 1 *S* from each condition per block.

All groups except one with no training received two paired-associate learning tasks. The stimuli in the training tasks were presented visually, while in the transfer task either the same or different stimuli, depending on the treatment condition, were presented via the tactual modality. This sequence was used exclusively, since there was no interest in the well-known asymmetry in visual–tactual vs. tactual–visual transfer (e.g., Gaydos, 1956). In addition, Lobb's (1970) finding of no transfer from touch to vision but reliable transfer from vision to touch in a discrimination–recognition task suggests that a visual-to-touch training–transfer sequence would optimize the transfer of stimulus learning. The treatment conditions were differentiated only by the conditions of training. The seven training

conditions are described below, after which the rationale is given for the inclusion of the various conditions.

Group 1: Six CMPs were paired with six motor switching responses in the first training task. The transfer task was essentially a continuation of training with the exception of the change in modality of the stimulus presentation. Thus, *S* first learned to associate a particular motor switching response to a particular visual CMP, and then learned to associate the analogue tactual form to the same motor response.

Group 2: Same as Group 1, except that the response terms in the training task consisted of six letters of the alphabet (A, E, H, N, O, and R). The six letters ranked within the top nine of the alphabet in absolute frequency of occurrence (Underwood & Schulz, 1960).

Group 3: Same as Group 1, except that a different set of six CMPs were used in training.

Group 4: Same as Group 3, except that the response terms in training were the same letters used for Group 2. This is the WC nonspecific control.

Group 5: Same as Group 1, except that the stimuli used in the training task were six CMHs, thus making the stimuli different from those in the transfer task.

Group 6: Same as Group 5, except that the response terms in training were the same letters used for Group 2. This is the BC nonspecific control.

Group 7: Same as Group 1, except that the training task was not administered.

The transfer of forward associations is given by contrasting the transfer performance of Groups 1 and 2 under the assumption that superior transfer for Group 1 is not due to the transfer of response learning or, alternatively, is not spuriously given by the depressed transfer performance of Group 2 due to the transfer of inappropriate response terms. Violation of either assumption would give an inappropriate estimate of the transfer of forward association. The first assumption is readily met by trivializing response learning in the training task by using single-letter response terms and making them available for continual inspection. The second assumption is not as easily met; however, this source of forward transfer can be minimized by using training- and transfer-task response terms from grossly different categories (Ellis, 1969, p. 408).

The transfer of stimulus learning is provided by a direct comparison between Groups 2 and 4. Both groups contain the operations for potential sources of WC nonspecific transfer; however, since Group 2 has the same nominal stimuli in both tasks, the conditions are provided for the transfer of stimulus learning. Accordingly, if Group 2 learns some *specific* stimulus encoding skills associated with the visual training stimuli and if these skills transfer to the tactual transfer task when the same nominal stimuli are presented, then Group 2 will evidence superior transfer performance relative to the nonspecific control (Group 4).

Reliably more transfer for G[illegible]ompared to Group 6 would indicate that n[illegible] transfer in an intersensory transfer paradi[illegible]pendent on the class of form stimuli that are employed. A comparison of Groups 6 and 7 would yield evidence as to whether any nonspecific effects are [illegible]lent.

The inclusion of Groups 3 an[illegible]plements of Groups 4 and 6, respectively, w[illegible]rovide evidence as to whether any reliable t[illegible] response learning was present, since t[illegible]er groups' response terms were used in both tasks; hence, equivalent transfer performance for both "pairs" would further substantiate the above assumption that sources of transfer due to response learning were effectively eliminated.

Visual training. A standard paired-associate anticipation learning situation was presented to all *S*s except the no-training control condition. The 6-sec. intrastimulus interval included [illegible] stimulus exposure and a 3-sec. [illegible] response exposure.

Each of the six forms was presented for 10 anticipation trials for all training conditions. Evidence obtained from pilot data indicated that asymptotic performance would be achieved with this number of trials. An 8-sec. interstimulus (ISI) and intertrial interval (ITI) were employed. The ISI and ITI were kept the same so as to minimize the possibility that *S* could exclude certain response alternatives (switches or letters) as having already occurred within any one trial block. The stimulus–response pairs were randomized within each trial with two restrictions: (*a*) no more than two consecutive switching responses in either ascending or descending series were allowed, e.g., Switches 1, 2, and 3 could not occur as response terms for the first three forms; and (*b*) the same stimulus–response pair could not occur as the last pair of one trial and the first pair of a following trial.

Standard paired-associate instructions were read to all *S*s receiving training. The six letters were printed on a card and given to *S*s using the letters as response terms. The instructions were that *S*s could refer to the card as often as needed, but that they were to rely on the card as little as possible.

Tactual transfer. All *S*s received the same instructions and the same tactual paired-associate learning task. For each stimulus–response presentation, *S*s were instructed, following a ready signal from *E*, to locate with their index finger a small notch centered on the front edge of the plastic mount. When this was done, *S*s then located the attached form by pushing the index finger forward. The forms retained the same orientation trial-to-trial within the holders. Moreover, for Groups 1 and 2, the placement of the forms for tactual manipulation was such that the top of the forms when presented visually coincided with the most distant aspects of the forms when presented tactually. The 14-sec. ISI, including a 7-sec. anticipation interval and a 7-sec. stimulus and response interval, was started when *S*s began tracing the perimeter of the forms with the index finger. All tracing was done in a

clockwi... ...ction, and *Ss* were instructed to pace themse... ...o that at least one complete tracing of the fo... ...ould be achieved before the correct response ... s displayed. Upon presentation of the corr... response, as indicated by a light located abov... the correct switch, *Ss* were instructed to continu... tracing the form until the "correct green light goes out." Again, as was the case with the visual tr...ing, the ISI and ITI were kept constant at ...roximately 8 sec.

The six forms were presented for 10 anticipation trials following 1 study trial. The same "randomiza... ...rocedure used in the visual training task was also used... The *Ss* were not informed as to ...hether the same or different forms would be pre... ...d in the transfer task. If *S* specifically asked ... information, he was told that a study trial ... be given first, and he could determine this ...imself.

Results

Training. An analysis of the number of correct responses for Groups 1 through 6 across five blocks of two trials each did not yield a reliable difference, F (5, 78) = .40, nor was the Groups × Trials effect reliable, F (20, 312) = .73. Both findings are commensurate with the assumption that the six experimental groups were equivalent with respect to their learning rates during acquisition, hence differences in transfer performance must reflect paradigmatic operations and not differences in degree of original learning.

Transfer. Table 1 presents the transfer results in the form of mean correct responses averaged across all 10 transfer trials for each of the seven groups.

The comparison of initial interest, although itself nonanalytical in character, was that of Groups 1 and 7. A difference in favor of Group 1 has typically been taken as evidence for the transfer of specific associative habits. It was necessary that this difference be shown to be reliable, since all of the separate component sources of transfer were embodied in Group 1. An application of Dunnett's test to this comparison revealed a reliable difference, t (91) = 3.15, $p < .01$.

Evidence for the transfer of forward associations is given by comparing the performance of Groups 1 and 2. An analysis of the transfer performance for these two groups yielded a significant difference, t (26) = 2.22, $p < .05$. That the superior transfer performance obtained by Group 1 is not in any way attributable to the transfer of response learning is indicated by identical performances of Groups 5 and 6, and nearly equal performances of Groups 3 and 4, $t(26) < 1$.

TABLE 1

Mean Correct Responses Averaged Across All 10 Transfer Trials for Each Paradigm

Group	Mean correct responses	*SD*
1	33.64	8.58
2	26.21	9.14
3	23.64	9.77
4	23.14	5.54
5	25.78	9.38
6	25.78	7.05
7	23.28	10.37

Note. See text for description of groups.

Although Group 2 did evidence more correct transfer responses than any other control group (3 through 7), as indicated in Table 1, the Dunnett test contrasting Group 2 with 7 did not yield a reliable difference, t (91) < 1. Accordingly, none of the other control groups (3 through 6) were reliably superior to the no-training control; and the combined Groups 3 and 4 when compared to the combined Groups 5 and 6 did not yield a significant difference, t (54) = 1.12. No reliable evidence was therefore obtained in the present study for the transfer of any nonspecific sources.

Discussion

The principal finding of the present study was the positive transfer of forward associations evidenced by Group 1, which received the same stimuli and responses in both the training and transfer task. This finding is taken as an unequivocal demonstration of the transfer of forward associations under conditions where the mode of stimulus input is changed.

Two contrasting interpretations of this finding are possible. First, it could be assumed that *Ss* had, in advance of the present study, established common mediating responses to the intersensory equivalents of the six stimuli, or fractions thereof, and that the transfer of newly learned associations was therefore made possible via these mediators. A four-stage stimulus

equivalence paradigm, A–B, C–B, A–D, C–D, would provide an account consistent with the observed transfer of forward associations. It should be noted that this interpretation makes no necessary assumptions about whether the A and C terms possess any common attributes. The A and C terms could become functionally equivalent due simply to the addition of common mediating responses.

An alternative interpretation (Gibson, 1966, p. 55) is that although the stimulus objects are presented to different sensory systems, *S*s can learn to encode on features of the input that are invariant with the sensory system to which they are presented. The degree of intersensory transfer would be directly related to *S*'s ability to respond to features of information that are common to the various input systems.

The procedure for generating form stimuli in the present study was chosen because the forms were not highly representative of stimuli with which *S*s would have had previous experience. Accordingly, the likelihood of the transfer of forward associations being attributable to previously learned mediators would be minimal. On the other hand, acceptance of the second explanation given above as a basis for the transfer of forward associations minimally requires that it be shown that *S*s can in fact encode stimulus attributes that are invariant to the mode of input. Group 2 provides the *conditions* for satisfying this assumption. Recall that the only source of positive transfer other than nonspecific sources for Group 2 was the transfer of stimulus learning. In short, if *S*s in the training task learned specific stimulus encoding skills for the training stimuli, and if these skills transferred to the situation where the same stimuli were presented via another sensory system, then it must perforce follow that *S*s were responding to some common stimulus attributes in both tasks. Although Group 2 showed more transfer than the nonspecific controls, this effect was not reliable. Therefore, the results of the present study, although suggesting support for a "higher level" encoding system insensitive to mode of sensory input, do not converge on either alternative explanation. Ellis (1973) has, however, shown the transfer of stimulus learning in an intramodal transfer paradigm with 24-point shapes as stimuli, and has proposed that the transfer of stimulus learning depends on complex stimuli containing multiple redundant cues. Perhaps support for the invariant-features hypothesis could be obtained by employing more complex stimuli in [illegible]tersensory transfer paradigm, thereby p[illegible]g the basis for possible stimulus learning [illegible]r.

The reasons for the failu[illegible]tain any transfer attributable to non[illegible]sources of facilitation, as evidenced by [illegible]t that the transfer performance of Grou[illegible]hrough 6 was comparable to Group 7, a[illegible]ot immediately apparent. This finding, [illegible]ver, is not unique, as Clark, Warm and, Se[illegible]sky (1972) also recently reported no evi[illegible]e for nonspecific transfer in an inters[illegible] transfer paradigm.

Several possiblities exist that [illegible], in part, explain the lack of nonspecific t[illegible]fer in the present study. First, the tra[illegible], criterion was fixed at 10 trials for all gro[illegible]since preliminary evidence from pilot inves[illegible]tions indicated that near asymptotic perfo[illegible]ce would be obtained. Schwenn and Po[illegible] (1967) have found a direct relationship [illegible]ween the level of training and the amount [illegible]nspecific transfer. The possibility exists[illegible]herefore, that nonspecific sources of transfe[illegible]ld have been operative in the present st[illegible] had the level of training been extended [illegible]nd the criterion of 10 trials. Second, re[illegible]vidence (Rogers, 1968) suggests that a pa[illegible] what is learned as a nonspecific component i[illegible] rhythm of observing stimuli and a latency [illegible] giving overt responses. Rogers found t[illegible] decrements were observed when the a[illegible]ipation intervals were shortened or lengthened. He argued that a reduction in performance correlated with a change in the intervals in the direction of making them longer might be due, in part, to a disruption of stable pacing behavior which in turn would affect specific techniques in learning to make associations. Therefore, the change in intervals from the training to the transfer task in the present study might have contributed to the failure to obtain any nonspecific transfer. The fact that Holmgren et al. (1966) found evidence for BC nonspecific transfer in an intersensory transfer paradigm with the use of consistent training-transfer anticipation intervals would support this interpretation.

Both of these interpretations of the lack of nonspecific transfer address themselves to training-transfer parameters, variables unrelated to changes in stimulus modality. Alternatively, it could be argued that part of the nonspecific skills that are learned is how to encode the information for particular modes on input, implying, therefore, no transfer of those skills when the stimulus modality is changed.

REFERENCES

BJORK, M., GARVILL, J., & MOLANDER, B. Cross-modal transfer and distinctiveness of stimulus objects. *Reports from the Psychological Laboratory, University of Stockholm*, 1965, No. 186.

CL..., J. L., WARM, J. S., & SCHUMSKY, D. A. General and specific factors in the intersensory transfer of form. *Journal of Experimental Psychology*, 1972, **95**, 184–188.

...IS, H. C. Transfer and retention. In M. H. Marx (Ed.), *Learning Processes*. New York: Macmillan, 1969.

ELLIS, H. C. Stimulus encoding processes in human learning and memory. In G. H. Bower (Ed.), *[illegible] of learning and motivation*, Vol. 7. New York: Academic Press, 1973.

...S, H. F. Intersensory transfer in the discrimination of form. *American Journal of Psychology*, 1956, **69**, 107–110.

...N, E. J. *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts, 1969.

GIBSON, J. J. *The senses considered as a perceptual system*. Boston: Houghton, 1966.

HOLMGREN, G. L., ARNOULT, M. D., & MANNING, W. H. Intermodal transfer in a paired-associate learning task. *Journal of Experimental Psychology*, 1966, **71**, 254–259.

Lobb, H. Asymmetrical transfer of form discrimination across sensory modalities in human adults. *Journal of Experimental Psychology*, 1970, **86**, 350–354.

POSTMAN, L., & SCHWARTZ, M. Studies of learning to learn: I. Transfer as a function of method or practice and class of verbal materials. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 37–49.

ROGERS, J. G. Input and output speed components of learning-to-learn. *Journal of Experimental Psychology*, 1968, **78**, 233–241.

SCHWENN, E., & POSTMAN, L. Studies of learning to learn: V. Gains in performance as a function of warm-up and associative practice. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 565–573.

SINHA, A. K. P., & SINHA, S. N. Intersensory transfer in learning sequences. *Journal of Experimental Psychology*, 1960, **60**, 180–182.

THURMOND, J. B., & ALLUISI, E. A. An extension of the information-deductive analysis of form. *Psychonomic Science*, 1967, **7**, 157–158.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

WRIGHT, J. M. Cross-modal transfer and sensory equivalence—A review. *Scandinavian Journal of Psychology*, 1970, **11**, 22–30.

(Received June 11, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 954-962

PARTIAL REWARD EITHER FOLLOWING OR PRECEDING CONSISTENT REWARD: A CASE OF REINFORCEMENT LEVEL[1]

E. J. CAPALDI[2]

Purdue University

Previous investigations in which partial reward either preceded consistent reward (P-C) or followed it (C-P) employed the same reward magnitude in connection with both schedules. Findings similar to those previously obtained were obtained here when the reward magnitude associated with the C schedule was larger than that associated with the P schedule. However, associating smaller reward with the C than with the P schedule reduced resistance to extinction under P-C conditions and increased resistance to extinction under C-P conditions. These and earlier separate-phase extinction findings were found to be compatible with the principle of reinforcement level, which also has implications for positive and negative contrast effects, as well as a variety of other phenomena such as the frustration suppression effect.

In separate-phase investigations, a particular level of some variable occurs for a block of trials, e.g., large reward, partial reward, etc., followed in a second block of trials by a different level of the variable, e.g., small reward, consistent reward, etc. Separate-phase investigations have been concerned primarily either with early postshift behaviors, i.e., positive and negative contrast effects, or with extinction behavior, or with both. Of these findings, only negative contrast is explained reasonably well by available hypotheses (see Capaldi, 1972). Most hypotheses do not attempt to explain positive contrast (see Weinstock, 1971), and a variety of separate-phase extinction findings either contradict or are not explained by available hypotheses (e.g., Robbins, Chait, & Weinstock, 1968; Spear, Hill, & O'Sullivan, 1965; Sutherland & Mackintosh, 1971). The purpose of the present investigation was to test a hypothesis, to be considered later, which is consistent with findings reported in available varieties of separate-phase investigations. The basic idea contained in this hypothesis is that any change in reward schedule affects reinforcement level, which determines stimulus control over behavior and thus performance.

In previous investigations partial reward has both preceded consistent reward (P-C) and followed it (C-P). C-P training consisting of 100 C and 60 P trials reduced resistance to extinction very substantially relative to 60 P trials (Sutherland, Mackintosh, & Wolf, 1965; see Sutherland & Mackintosh, 1971, for other similar findings). Under P-C conditions a slight decrease in resistance to extinction was associated with a substantial number of interpolated C trials (70 P and 70 C trials) but not with a moderate number (70 P and 25 C trials) in an investigation by Theios (1962; see also Sutherland et al., 1965). Sutherland and Mackintosh have reviewed this literature and have characterized its findings as follows: The first of the two schedules given has had a major or disproportionate effect on resistance to extinction. This means that if enough first-schedule training is provided, extinction will be rapid if the first schedule is C and slow if the first schedule is P. Sutherland and Mackintosh have suggested that this disproportionate first-schedule effect, especially for C-P, while consistent with their attention model, is not consistent with hypotheses more commonly applied

[1] This research was supported in part by National Institute of Child Health Development Grant 5R-HD-04379-03 to the author.

[2] Requests for reprints should be sent to E. J. Capaldi, Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907.

to re schedule effects such as the frust (Amsel, 1958) and sequential (Cap , 1966) hypotheses.

Th sequential hypothesis explains reward schedule and other effects by emphasizing the control over behavior exercised by both external cues, e.g., alley cues, and internal cues, e.g., those produced by nonreward, smaller reward, and so on. This sort of analysis might be successfully applied to P-C and C-P training, it was hypothesized, by assuming that changes in reward percentage involve the same sorts of principles as changes in reward magnitude. Within the sequential hypothesis, it has been assumed that the capacity of available stimuli to evoke the reaction is an increasing function of reward magnitude (e.g., Capaldi & Capaldi, 1970; Leonard, 1969).

In available P-C and C-P investigations, the same reward magnitude has been associated with the P and the C schedules. In the present investigation, the reward magnitude associated with the C schedule (either 1 or 12 .045-gm. pellets) was either less than or greater than that associated with the P schedule (always 6 pellets). The five groups employed were designated C1-P6, C12-P6, P6-C1, P6-C12, and P6.

Method

Subjects. The *S*s, about 120 days old at the start of the experiment, were 50 naive male rats purchased from the Holtzman Co., Madison, Wisconsin.

Apparatus. The apparatus was a straight alley, 208.3-cm. long, 10.2-cm. wide, and enclosed by 22.9-cm.-high sides, covered with hinged 1.3-cm. hardware cloth. It was constructed of wood and painted a medium gray. It had three basic sections: a 35.6-cm. start section, a 117.1-cm. run section, and a 40.6-cm. goal section. On all trials, *S* was placed at the rear of a 25.4-cm. start treadle suspended over a microswitch, starting a .01-sec. clock. This clock stopped and a second started when *S* broke a photobeam 10.2 cm. from the treadle's tip (start time). The interruption of a second beam 117.1 cm. from the first stopped the second clock (run time) and started a third clock. The third photobeam was 30.5 cm. from the second beam and 5.1 cm. from the front edge of a brass 5.1 × 10.2 × 3.8 cm. food cup. Interrupting it stopped the third clock (goal time). When *S* broke the third beam a brass guillotine door, 30.5 cm. from the alley's distal end, was lowered manually, thereby confining *S* to the goal area. Individual times for all three sections were recorded, as well as "total" time, which was the sum of the elapsed times on the clocks.

Procedure. The *S*s were individually caged, fed 12 gm. of Lab Blox each day, water always being available, and on Days 7–10 handled in groups of three for about 3 min. On Days 8–10 each *S* received six .045-gm. Noyes pellets in the home cage.

On Day 11, which began experimental training, *S*s were randomly divided into five groups of 10 *S*s each. In each of the three phases of experimental training there were five trials each day. Phases 1 and 2 lasted for nine days each and Phase 3 (extinction) for eight days. Reward consisted of .045-gm. pellets, nonreward in all phases of 20-sec. confinement in the unbaited goalbox. Group C1-P6 received one pellet consistent reward in Phase 1 and six pellets partial reward in Phase 2. The remaining groups were designated in identical fashion. Group C1-P6 along with Groups P6-C1, C12-P6, and P6-C12 comprised the factorial portion of the experimental design. Group P6 began training in Phase 2. Three schedules of partial reward were used in rotation over the nine days of each acquisition phase, RNRNR, RNNRR, and RNNNR, where R denotes six-pellet reward and N, nonreward.

The *S*s were brought into the room in squads of five in holding cages, a squad consisting of one *S* from each group. The *S*s were run in rotation, running order within a squad being varied randomly each day. In Phase 1 Group P6 *S*s were handled briefly during their order in the rotation. The intertrial interval was about 5–8 min. in each phase. Following the last daily trial, *S*s were returned to the home cage, where the 12-gm. daily ration was provided about 15 min. later.

Results

Time scores in each alley section were converted to speeds. The measure employed was the daily median speed for each *S* in each alley section.

The performance of each of the groups in each of the alley sections on the last two days of Phase 1 training (P1) and on each of the training days in Phase 2 is shown in Figure 1. An analysis on the last two days of Phase 1 training indicated that differences were significant in each of the alley sections, start; $F(4, 45) = 4.31$, $p < .01$; run, $F(4, 45) = 4.06$, $p < .01$; and goal, $F(4, 45) = 3.98$, $p < .01$. Newman–Keuls tests indicated that in start and run Group C1-P6 ran more slowly than each of the remaining groups ($p <$

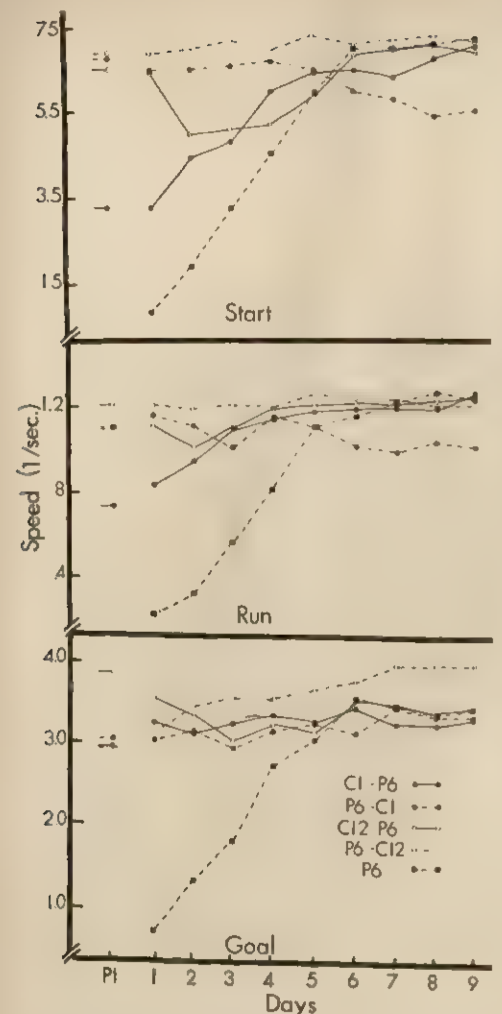


FIGURE 1. Median daily speeds for each of the groups in each of the alley sections of the last two days of Phase 1 training and on each of the training days in Phase 2.

.01), and in goal Group C12-P6 ran more rapidly than each of the remaining groups ($p < .01$), no other differences being significant. It is sometimes reported (e.g., Goodrich, 1958) that when partial and consistent groups receive the same reward magnitude, the partial group is faster than the consistent group in the start and run sections of the alley but slower than the consistent group in the goal section of the alley. Thus, it is unclear whether the superiority of Groups P6-C1 and P6-C12 relative to Group C1-P6 in start and run was a reward magnitude eff… reward percentage effect, or a combin… of both.

On the last two days of Ph… training differences were significant in …h of the alley sections; start, F (4… 3.09, $p < .05$; run, F (4, 45) = 2… $< .05$; and goal, F (4, 45) = 2.81, p … New-man–Keuls tests indicated … start and run Group P6-C1 ran … slowly than each of the remaining … ($p <$ .05), and in goal Group P6-… n more rapidly than the remaining … ($p <$ .05), no other differences bein… nificant. Thus, terminal Phase 2 differ…s as a function of reward condition w… much like terminal Phase 1 difference… func-tion of reward condition, e… that terminal differences were smaller… Phase 2 than in Phase 1.

Extinction. In view of the …rminal acquisition differences obtaine… ere, in some cases it is best to em… ize in extinction the interaction of … icular variables with extinction days. …he ex-tinction data were analyzed in … ways. Following a one-way analysis of …riance, a 2 × 2 factorial analysis was p…formed (Group P6 omitted) employing t… error term from the one-way analy…. The factorial analysis evaluates the effects of reward magnitude (C12 vs. C1) and order of presentation of P and C (P-C vs. C-P).

The performance of the five groups in each of the alley sections of each of the extinction days is shown in Figure 2. Figure 2 shows that Group P6-C1 tended to extinguish rapidly in early alley sections and that Groups C1-P6 and P6 tended to extinguish slowly in later alley sections. The one-way analysis indicated that the groups differed significantly in each of the alley sections; start, F (4, 45) = 6.71, $p < .001$; run, F (4, 45) = 9.48, $p < .001$; and goal, F (4, 45) = 6.81, $p < .001$, and that the Groups × Days interaction was significant in each of the alley sections; start, F (28, 315) = 1.75, $p < .01$; run, F (28, 315) = 2.08, $p < .01$; goal, F (28, 315) = 1.71, $p < .01$. Newman–Keuls tests indicated that Group P6-C1 ran slower than all other groups in start and in run (all ps $< .01$), that Group C1-P6

differ from all other groups in run (ps < .05 or better) and in goal ($p < .01$), and Group P6 differed from all other groups in goal ($ps < .01$), no other differences being significant. Except for Group P6-C1, these significant differences meaningful, for in no other case could they be due to inferior terminal acquisition performance.

In start, the effects of reward magnitude, F (1, 45) = 4.13, $p < .05$; of presentation order, F (1, 45) = 7.09, $p < .05$; and the Reward Magnitude × Presentation Order interaction, F (1, 45) = 10.91, $p < .01$, significant. In addition, each of these interacted significantly with days, Reward Magnitude × Days, F (7, 315) = 2.11, p < .05; Presentation Order × Days, F (7, 315) = 2.16, $p < .05$; and Reward Magnitude × Presentation Order × Days, F (7, 315) = 2.31, $p < .05$. The major basis of these significant effects in start was the rapid extinction over days of Group P6-C1. In start, then, C1 produced less resistance to extinction than C12 when combined with P6, and P-C produced less resistance to extinction than C-P. In goal, all these effects were, again, significant, reward magnitude, F (1,45) = 4.27, p < .05; presentation order, F (1, 45) = 4.13, $p < .05$; Reward Magnitude × Presentation Order, F (1, 45) = 5.27, $p < .05$; Reward Magnitude × Days, F (7, 315) = 2.26, $p < .05$; Presentation Order × Days, F (7, 315) = 2.09, $p < .05$; and Reward Magnitude × Presentation Order × Days, F (7, 315) = 2.45, $p < .05$. Unlike in start, the major basis of the goal effects was the great resistance to extinction of Group C1-P6. In goal, then, C1 produced greater resistance to extinction than C12 when combined with P6, and C-P produced greater resistance to extinction than P-C. In run, the only significant effects were those associated with Reward Magnitude × Presentation Order, F (1, 45) = 14.82, $p < .01$, and with Reward Magnitude × Presentation Order × Days, F (7, 315) = 2.82, $p < .01$. The trend in run, then, was intermediate to the extremes shown for start and goal, thus producing significant differences only

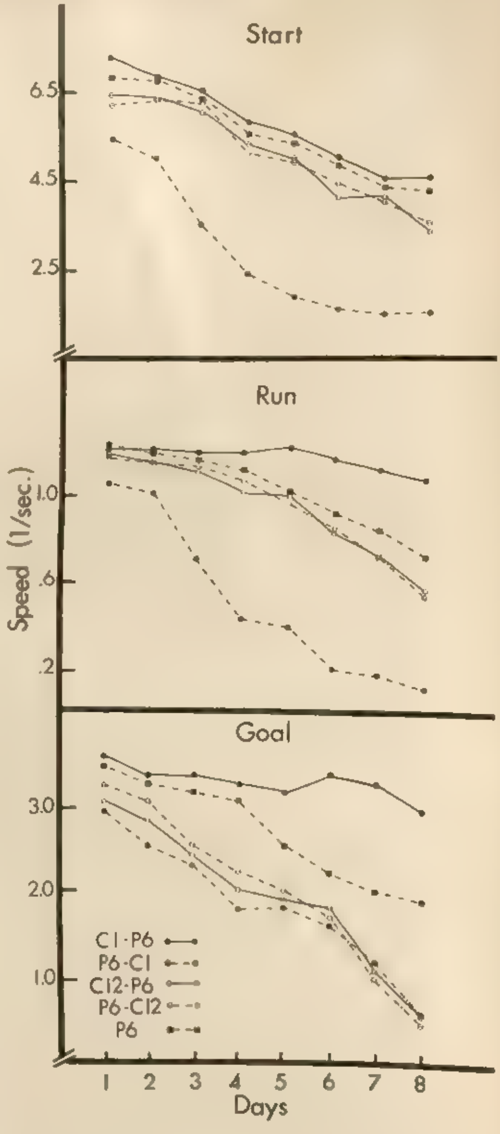


FIGURE 2. Median daily speeds for each of the five groups in each of the alley sections on each of the extinction days.

when reward magnitude and presentation order interacted. The major findings that emerged in extinction were as follows. Group C1-P6 showed very great resistance to extinction relative to all other groups in the goal section, a tendency which, while somewhat evident in the run section, was evident only in relation to Group P6-C1 in the start section. In each alley section, Group P6 performed much as did

Group C1-P6 except, of course, that Group P6 ran more slowly than Group C1-P6 in the run and goal sections. Group P6-C1 extinguished rapidly relative to all other groups in the start section, a tendency which, while somewhat evident in the run section, was evident only in relation to Groups C1-P6 and P6 in the goal section. Finally, Groups P6-C12 and C12-P6 failed to differ from each other in any alley section.

Discussion

In the runway, small consistent reward has produced greater resistance to extinction than large consistent reward and less resistance to extinction than a larger magnitude of partial reward (e.g., Hulse, 1958). However, whatever might have been the effects on extinction of C1 in isolation, these cannot have been its effects when employed in combination with P6, for Group C1-P6 was very resistant to extinction, while Group P6-C1 extinguished very rapidly. Such interactive effects between schedules, while new to C-P and P-C investigations, are not new to separate-phase investigations. The major examples of such interactive effects, of course, are the positive contract effect, or better performance on the part of a group shifted from small to large reward (Group C1-P6 was shifted from small to large reward) than on the part of a large-reward control, and the negative contrast effect, or poorer performance on the part of a group shifted from large to small reward (Group P6-C1 was shifted from large to small reward) than on the part of a small-reward control group. The problem for theory is to identify the processes responsible for these interactive effects. The attention theory of Sutherland and Mackintosh (1971) is so stated that it expects that C-P training will either reduce resistance to extinction or at least produce no greater resistance to extinction than P training (the precise form of the deduction depends upon number of P training trials). Also, given that Group C1-P6 showed very great resistance to extinction, the Sutherland and Mackintosh model would not expect Group P6-C1 to extinguish rapidly (see Sutherland & Mackintosh, 1971). However, inability to deal with the Reward Magnitude × Presentation Order interaction obtained here is not confined to attention theory. It is suggested that the present Reward Magnitude × Presentation Order interaction is simply an example of interactive effects between reward schedules, which at present are not well understood. A model is considered below that appears to adequately explain known interactive schedule effects, whether manifested in extinction, negative contrast, or positive contrast.

Traditionally, the conditioning of stimuli to responses has been considered to be accomplished through either contiguity or reinforcement. The sequential view (e.g., Capaldi, 1966), for example, emphasizes reinforcement. The reinforcement assumption of the sequential model is replaced by that of reinforcement level, all other assumptions of that view being retained. It is suggested that strength of conditioning is always determined by the relationship between the reward that is expected and the reward that is obtained. This relationship determines reinforcement level. More specifically, the view that strength of conditioning is determined by reward magnitude defined physically in terms of weight of incentive, etc. (e.g., Capaldi, 1966), is replaced by the view that strength of conditioning is determined by reinforcement level (obtained reward is, of course, defined physically). Accordingly, if obtained reward is greater than expected reward, then available stimuli will acquire a greater capacity than otherwise to elicit the reaction. However, if obtained reward is smaller than expected reward, then available stimuli will suffer a decrease in their capacity to elicit the reaction (unconditioning). Stimuli that are unconditioned acquire an increased capacity to elicit competing reactions. Variables that are assumed to determine expected reward are considered below, along with selected assumptions of the sequential view that are relevant to understanding reward schedule effects.

Expected reward is assumed to increase as reward magnitude increases and to increase as reward percentage increases. Accordingly, with reward magnitude held constant, expected reward will be higher the greater the reward percentage. Expected reward is assumed to increase along with number of training trials. Under consistent reward, conditioning will eventually be weak, because under consistent reward the value of expected reward eventually becomes identical to that of obtained reward. Under partial reward, the value of expected reward eventually stabilizes between two values of obtained

reward, that available on rewarded trials and that available on nonrewarded trials. Under partial reward, then, there is both strong conditioning (rewarded trials) and unconditioning (nonrewarded trials). Adjustment of expected reward to prevailing reward condition will be fastest where no specific former expected reward level has been established (as in Group P6), and it will be slower the stronger is the former expected reward (as in Group C1-P6). In considering groups such as P6 vs. C1-P6, or groups that may be currently receiving the same reward magnitude but that have different expected reward values, it should be borne in mind that not only are rewarded-trial conditioning effects stronger where expected reward is lower, but nonrewarded-trial unconditioning effects are weaker. Accordingly, in comparing groups of this sort below, only rewarded-trial conditioning effects will be considered in the interest of simplicity. It has been suggested (e.g., Capaldi, 1972) that nonreward or smaller reward produces an internal stimulus (a memory) which has two components, an emotional or aversive component, and a component that corresponds to such characteristics of the reward as its size and so on (perceptual or cognitive component). Call this stimulus S^N. The frustrative or aversive components of S^N occur only when obtained reward is below expected reward, the perceptual components of S^N being independent of expected reward. Since S^N is a memory, the frustrative components of S^N are available to be conditioned or unconditioned to the reaction only on trials which follow trials on which obtained reward was below expected reward. These assumptions are now applied to present and previous separate-phase findings.

When the same reward magnitude is associated with the P and C schedules and P training follows C training, the higher expected reward level appropriate to the C schedule persists for some number of P trials, and thus, on those P trials S^N is less strongly conditioned to the instrumental reaction than where P-alone training is provided, and thus, C-P is less resistant to extinction than P (see Sutherland & Mackintosh, 1971). Under same-magnitude P-C conditions, expected reward level rises over the C trials, which would tend to produce less resistance to extinction in P-C than in P, but at the same time, a variety of stimuli gain in their capacity to evoke the reaction over the C trials (expected reward is below obtained reward on initial C trials). This would tend to oppose the greater eventual unconditioning in P-C relative to P, and perhaps accounts for the fact that, even following extensive interpolated C training, P-C is not much less resistant to extinction than P (Theios, 1962). Employing the reasoning used above, it can be understood why in the present investigation Groups C12-P6 and P6-C12 were less resistant to extinction than Group P6.[3]

In dealing with the case in which the C schedule reward magnitude is smaller than the P schedule reward magnitude, the specific assumption is entertained that the expected reward level associated with P6 was higher than that associated with C1. For Group P6-C1, then, on some number of C1 trials expected reward was above obtained reward (unconditioning), and therefore S^N lost some of the capacity to evoke the reaction that it had acquired under P6 training. For Group P6-C1, then, at least some C1 trials functioned as nonrewarded extinction trials (although less so), and thus, Group P6-C1 was less resistant to extinction than Group P6 and Group P6-C12. For Group C1-P6, on the other hand, expected reward level was below obtained reward for some number of P6 trials (strong conditioning of S^N), and thus, Group C1-P6 was very resistant to extinction. Of course, expected reward was below obtained reward in Group P6 as well, but in Group P6 expected reward was not below obtained reward for as many trials as in Group C1-P6 (previously established expected reward levels are resistant to change).

The considerable similarity between the frustration suppression effect, or FSE (e.g., Dunham & Marx, 1967), and particularly Groups C12-P6 and P6-C1 here is worth noting. A group that receives a larger reward each day in magazine training than in lever

[3] After this report was prepared, an investigation by Mellgren, Lombardo, Wrather, and Weiss (1973) appeared which showed that C-P (20 trials in each phase) produced greater resistance to extinction than P (20 trials) when the same reward magnitude was associated with each schedule. Following only 20 C trials, expected reward would not be firmly established, and thus, S^N was conditioned to the reaction in the P phase about as strongly in Group C-P as in Group P. Too, where only 20 P trials are employed, alley and other stimuli would not be strongly conditioned to the reaction in Group P. Such stimuli would be more strongly conditioned to the reaction in Group C-P due to the additional 20 C trials, and thus, Group C-P was more resistant to extinction than Group P.

training (downshifted group) is less resistant to extinction than a control that receives the same smaller reward magnitude under both magazine and lever conditions. It is reasonable to assume that the downshifted group is frustrated during lever training. As Dunham and Marx point out, a contiguity emphasis (e.g., Amsel, 1958) would suggest that the frustrative stimuli would acquire control over the instrumental reaction in the downshifted group and, therefore, elevate resistance to extinction in that group. The same, of course, would be expected on the basis of a traditional reinforcement view (e.g., Capaldi, 1966). On the other hand, the FSE is clearly consistent with the reinforcement level view. That is, the larger reward associated with magazine training would elevate expected reward in lever training, thereby allowing frustrative and other stimuli to acquire but little control over the instrumental reaction, and thus, the downshifted group would extinguish rapidly.

Other varieties of separate-phase extinction findings are consistent with the reinforcement level view, although these findings, like the FSE and C-P and P-C extinction findings, create considerable difficulty for available theories. It has been reported (e.g., Robbins et al., 1968; Spear et al., 1965) that groups given initial nonrewarded trials (no prior reward trials) prior to rewarded trials showed greater resistance to extinction than controls given rewarded trials only. These findings create difficulty for the sequential view, because only one transition from nonrewarded to rewarded trials occurs (see Capaldi, 1966), and they create difficulty for Amsel's (1958) frustration hypothesis, because initial nonrewarded trials are presumably not frustrative (see Robbins et al., 1964; Spear et al., 1965). According to the reinforcement level view, however, stimuli occurring in the rewarded phase would be more strongly conditioned to the instrumental reaction in the initially nonrewarded group (expected reward is below obtained reward) than in the rewarded-trial control (expected reward and obtained reward more quickly become the same). The similar finding reported by Capaldi (1970), that a small magnitude of consistent reward followed by a large magnitude of consistent reward produced greater resistance to extinction than large consistent reward alone, is also consistent with the reinforcement level view. The similarity of the above finding to those obtained here for Group C1-P6 is perhaps obvious; in all cases, postshift expected reward was below obtained reward. [illegible]ll cases, resistance to extinction was in[illegible] by postshift training. A difference be[illegible] the case being considered and the sa[illegible]agnitudes P-C vs. P case is worth no[illegible] to avoid possible confusion. Expected [illegible]ard level and thus unconditioning is gr[illegible] for P-C than for P in extinction. But [illegible] expected reward level of, e.g., an initially [illegible]rewarded group, is either below that or [illegible] to that of its rewarded control at th[illegible]e of extinction, depending on rewarde[illegible] training level.

When expected reward is ab[illegible]btained reward, both unconditioning an[illegible]^N occur. In most available negative contra[illegible]vestigations, S^N has not been conditi[illegible] to the instrumental reaction in preshift [illegible], most negative contrast findings can be sa[illegible]actorily explained in terms of the genera[illegible]ion decrement and other effects of S^N, w[illegible] appeal to unconditioning being unnece[illegible]y (see Capaldi, 1972). There appears to [illegible]t least one case of negative contrast, h[illegible]er, in which appeal to unconditioning as [illegible] as S^N is required. Ison, Glass, and L[illegible] (1969) reported that a shift from a large [illegible] small magnitude of partial reward produc[illegible] negative contrast effect, however, not [illegible]arly in postshift, as is usual, but only follo[illegible]ng considerable training. Early in pos[illegible]hift S^N would elicit responding strongly, [illegible]ause it was previously conditioned to the reaction by means of a large magnitude of partial reward. However, because in postshift expected reward was above obtained reward on both rewarded and nonrewarded trials, unconditioning would eventually be considerable, and presumably it was this rather substantial unconditioning effect that was responsible for the late-trial negative contrast effect reported by Ison et al. If the postshift condition following large partial reward is small consistent reward rather than small partial reward, the unconditioning effect would be less severe (no nonrewarded trials) and unconditioning would not occur over so many trials (the preshift and postshift expected reward level discrepancy is smaller). These are, perhaps, among the reasons why a shift from large partial reward to small consistent reward did not produce a negative contrast effect in an investigation reported by Capaldi and Ziff (1969).

Whether positive contrast has been obtained seldom or frequently seems, in part, definitional. If the meaning of positive contrast

is res[illegible] to cases in which reward magnitud[illegible]gher in postshift than in preshift, then [illegible]ve contrast has been obtained less frequ[illegible] than otherwise. According to the reinf[illegible]ment level view, however, the necessary [illegible] not sufficient conditions for positive con[illegible]st exist whenever expected reward lies be[illegible] obtained reward, which widens consi[illegible]rably the permissible instances of positive [illegible]trast. Thus, according to reinforcement level, positive contrast can occur on a shift from partial to consistent reward as well as on a shift from smaller to larger reward. [illegible]eung and Jensen (1969) reported positive [illegible]ntrast following shifts from various per[illegible]tages of partial to consistent reward. In [illegible] and run sections of the runway, [illegible]ial reward sometimes produces faster [illegible]ning than consistent reward (e.g., Good[illegible], 1959), which is, according to reinforce[illegible]nt level, a positive contrast effect. In the [illegible]al section, partial reward produces slower running than consistent reward, which is consistent rather than inconsistent with the present view. In the goal section, nonrewarded-trial unconditioning effects under partial reward are potent, and as Miller (1959) suggests (see also Capaldi & Levy, 1972), approach tendencies generalize more widely than inhibitory tendencies, which means that in earlier alley sections rewarded-trial conditioning effects under partial reward will be potent relative to nonrewarded-trial [illegible]conditioning effects. Positive contrast was [illegible]served by Robbins et al. (1968) on a shift [illegible]m initial nonrewarded trials to rewarded [illegible], although not classified as such by those authors. More orthodox instances of positive contrast were reported recently by Weinstock (1971) on a shift from small to large consistent reward, and by Mellgren (1971, 1972) on shifts from a small to a large magnitude of delayed reward.

Mellgren (1971) observed a positive relation between number of preshift trials and absolute speed on postshift trials, and this effect extended over 32 trials, at which point training was discontinued. According to the reinforcement level view, as the number of preshift trials increases, the number of postshift trials over which obtained reward will remain above expected reward increases. Thus, as number of preshift trials increases, the stronger should be postshift conditioning, producing, accordingly, a positive relation between number of preshift trials and absolute speed in postshift.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

Capaldi, E. D. Effect of an initial reward magnitude on subsequent resistance to extinction. *Journal of Experimental Psychology*, 1970, **86**, 283–287.

Capaldi, E. J. Partial reinforcement: A hypothesis of sequential effects. *Psychological Review*, 1966, **73**, 459–477.

Capaldi, E. J. Successive negative contrast effect: Intertrial interval, type of shift, and four sources of generalization decrement. *Journal of Experimental Psychology*, 1972, **96**, 433–438.

Capaldi, E. J., & Capaldi, E. D. Magnitude of partial reward: irregular reward schedules, and a 24-hr. ITI: A test of several hypotheses. *Journal of Comparative and Physiological Psychology*, 1970, **72**, 203–209.

Capaldi, E. J., & Levy, K. J. Stimulus control of punished reactions: Sequence of punishment trials and magnitude of reinforcement trials. *Learning and Motivation*, 1972, **3**, 1–19.

Capaldi, E. J., & Ziff, D. R. Schedule of partial reward and the negative contrast effect. *Journal of Comparative and Physiological Psychology*, 1969, **68**, 593–596.

Dunham, P. J., & Marx, M. H. Frustration-suppression effect in a discrete-trial operant situation. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 45–49.

Goodrich, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *Journal of Experimental Psychology*, 1959, **59**, 57–63.

Hulse, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *Journal of Experimental Psychology*, 1958, **56**, 48–57.

Ison, J. R., Glass, D. H., & Daly, H. B. Reward magnitude changes following differential conditioning and partial reinforcement. *Journal of Experimental Psychology*, 1969, **81**, 81–88.

Leonard, D. W. Amount and sequence of reward in partial and continuous reinforcement. *Journal of Comparative and Physiological Psychology*, 1969, **67**, 204–211.

Leung, C. M., & Jensen, G. D. Shifts in percentage of reinforcement viewed as changes in incentive. *Journal of Experimental Psychology*, 1968, **76**, 291–296.

Mellgren, R. Positive contrast in the rat as a function of number of preshift trials in the runway. *Journal of Comparative and Physiological Psychology*, 1971, **77**, 329–336.

Mellgren, R. Positive and negative contrast effects using delayed reinforcement. *Learning and Motivation*, 1972, **3**, 185–193.

Mellgren, R., Lombardo, J., Wrather, D., & Weiss, R. Partial reinforcement effect: the

expectancy of reward on nonrewarded trials. *Animal Learning and Behavior*, 1973, **1**, 105–108.

Miller, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill Book Company, 1959.

Robbins, D., Chait, H., & Weinstock, R. Effects of nonreinforcement on running behavior during acquisition, extinction, and reacquisition. *Journal of Comparative and Physiological Psychology*, 1968, **66**, 699–706.

Spear, N. E., Hill, W. F., & O'Sullivan, D. J. Acquisition and extinction after initial trials without reward. *Journal of Experimental Psychology*, 1965, **69**, 25–29.

Sutherland, N. S. [illegible] ackintosh, N. J. *Mechanisms of anima*[illegible]*mination learning*. N[illegible] York: Academic [illegible], 1971.

Sutherland, N. S[illegible]kintosh, N. J., & W[illegible] J. B. Extinct[illegible] a function of the order partial and co[illegible]t reinforcement. *Journal of Experimental* [illegible]gy, 1965, **69**, 56–59.

Theios, J. Th[illegible] reinforcement effect s tained through [illegible] of continuous reinforcement. *Journa*[illegible]*imental Psychology*, 19[illegible] **64**, 1–6.

Weinstock, R. [illegible]tion exploration of runway in the [illegible]ation of contrast eff[illegible] in the rat. *Jou*[illegible]*Comparative and Ph*[illegible] *logical Psycholog*[illegible]75, 107–115.

(Re[illegible]ly 3, 1973)

Journal of [illegible] Psychology
1974, Vol. [illegible], No. 6, 963–968

[illegible]D RECOGNITION AND MORPHEMIC STRUCTURE[1]

GRAHAM A. MURRELL[2] AND JOHN MORTON

Medical Research Council, Applied Psychology Unit, Cambridge, England

Six[illegible] *Ss* learned a w[illegible]d list and were then tested on recognition of words presented tachistoscopi[illegible]. When a test word was identical to one of the learned words, recognition was facilitated in the usual way. Pretraining with word[illegible]t was a different derivative of the same root morpheme as a test word [illegible] facilitation of recognition of that test word. However, learning [illegible] comparable visual-acoustic similarity but no morphemic relation [illegible] word gave only slight, nonsignificant facilitation of recognition. It [illegible]cluded that the process of recognizing a word involves assigning it to a [illegible]istic unit with specific semantic associations, i.e., a morpheme. Analysis [illegible]or responses suggested that, to some extent, root and suffix morphemes [illegible]re recognized independently, and that suffixes themselves may be subject to frequency effects. The principles underlying the phenomena studied may be most appropriately characterized as "the morpheme-frequency effect."

Words of higher frequency of occurrence in the language are more easily recognized than low-frequency words. The nature of the mechanism leading to this phenomenon is a subject of current debate (Broadbent, 1967, 1971; Catlin, 1969; Morton, 1968; Nakatani, 1970; Treisman, 1971). All current theories, however, are agreed that the recognition system is biased toward high-frequency words, although that bias is seen as operating in different ways. In the past, it was debated whether this bias was a function of the differential frequency of the perceptual experience of words (Howes, 1954; Neisser, 1954) or [illegible] frequency of production of words (D[illegible]ston, 1957), with corresponding definition of the unit of recognition as perception or response based.

[illegible] way of reconciling these positions was put forward by Morton (1964, 1969), who proposed the logogen model, in which the central units, logogens, were affected not only by visual and acoustic information, but also by semantic information. The latter is implicated in speech production and also in the recognition of connected material. Any logogen receiving more than a threshold amount of information "fires," making the corresponding word available as a response. As a result of firing, the threshold of the logogen is lowered and very gradually rises to a level slightly below its previous value. In this way, experience with words, both recognition and production, leads to permanent changes in threshold, biasing future responses.

The short-term change in threshold was included in the model to account for the finding that when a word is recognized, the same word will be recognized more easily if presented again within a short time.

The latter effect was utilized in the present experiment to create a short-term, experimentally controlled word-frequency effect. The technique has been used previously by Solomon and Postman (1952), who used nonsense words so there would be no "a priori" word-frequency effect. As will become clear, it was necessary to use familiar real words in the present experiment, but the uncontrolled a priori frequency effect could only mask the results and will not affect the conclusions.

Neisser (1954) found that pretraining with a word facilitated recognition of the same word but not of a homonym, e.g., after priming with "phrase," recognition of this word was easier than normal, but recognition of "frays" was not affected.

[1] This work was carried out at the Psychological Laboratory, University of Cambridge, while the first author was a final year undergraduate. The assistance of those who acted as *Ss* is gratefully acknowledged.

[2] Requests for reprints should be sent to Graham A. Murrell, Medical Research Council, Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, England.

TABLE 1

Experimental Words

(1)	(2)	(3)
Car	Cars	Card
Bored	Boring	Born
Sings	Singer	Singe
Pains	Pained	Paint
Lives	Lively	Livery
Sees	Seen	Seed
Man	Manned	Manner
Hanging	Hang	Hangar
Reader	Reading	Ready
Placed	Places	Placate
Walled	Wall	Wallet
Topped	Topping	Topple
Basis	Basic	Basin
New	Newer	Newt
Mile	Miles	Mild
Numbing	Numb	Numbers

His interpretation was that pretraining "does not merely raise the probabilities of certain verbal responses, but rather facilitates the perception of specific visual patterns [Neisser, 1954, p. 402]." Following this, Ross, Yarczower, and Williams (1956) attempted to demonstrate a monotonic relation between degree of facilitation and level of visual similarity between pretraining word and test word. They found no such relation, although they confirmed Neisser's main result. This, and other considerations, casts doubt upon Neisser's interpretation, and particularly his assumption that "phrase" and "frays" may be regarded as the same response (even assuming they are phonetically identical).

The interaction of semantic and sensory information necessary to account for the facilitative effects of a context on word perception (Morton, 1969; Rubenstein & Pollack, 1963; Tulving, Mandler, & Baumal, 1964) also necessitates a refinement of the definition of the unit of recognition. Thus, since the contexts appropriate to the two homographs "chop" are totally different, they might be expected to have functionally separate representations in the recognition system.

The question of interest in the present study is whether the unit concerned in recognition facilitation effects is a word as a visual–phonetic pattern, or a form bound to a specific semantic unit, i.e., a morpheme. The conclusions will bear on the nature of the response units made available by logogens.

The experiment investigated the effect on word recognition of pretraining with a word containing the same root morpheme, compared with pretraining with a word containing the same visual–phonetic pattern as its "root" when this is not actually the same morpheme. The effects of pretraining with the identical word and zero pretraining, as control conditions, were also tested.

Method

Subjects. Sixteen male undergraduates volunteered as *S*s, and were divided into two groups of 8. None were psychology students. Each completed the procedure once only, with no prior information about the experiment apart from the specific instructions (see below).

Materials. The apparatus was a Cambridge tachistoscope with an auxiliary timing device to permit control of flash duration in 5-msec. steps between 10 msec. and 140 msec. A neutral density filter (×16) was fitted over the viewing aperture to reduce stimulus contrast. The stimulus words were typed in lowercase on plain white cards, aligned so as to appear in the space between two short horizontal lines drawn on the "background" card upon which a word was superimposed in the stimulus flash. The *S* was provided with the tachistoscope trigger button on a cable, to initiate the flashes.

The experimental words are shown in Table 1. The three words in each row of the table have an initial sequence of letters in common. Group 1 and 2 words differ only in the inflection of their common root morpheme, but the same letter sequence in the Group 3 word is not the same morpheme, and the word bears no semantic relation to the other two. However, viewed purely as visual–acoustic patterns, the degrees of similarity between the words in different columns are comparable (e.g., the mean suffix length of Group 3 is intermediate between Groups 1 and 2, the difference between the means of Groups 1 and 2 being less than half of one letter).

Twenty other words were used in initial practice on tachistoscopic recognition. These were selected for low similarity with experimental words.

There were four sets of 12 words for pretraining. Each set comprised 4 words from each of Groups 1, 2, and 3, typed on a card in random order. Each *S* received a recognition test on the words of either Group 1 or Group 2 of Table 1. Thus, of the 16 words each *S* was tested on, 4 were subject to each of the four pretraining conditions, which were as follows: (*a*) pretraining with the same word (Condition S), (*b*) pretraining with a different inflection

of the same root morpheme (M), (*c*) pretraining with an unrelated word containing the same visual–acoustic form but different terminal letters (V), and (*d*) zero pretraining, no similar word having been presented (O).

Over the eight combinations of pretraining and testing word sets, every test word was subjected to every pretraining condition. Thus from each group of eight *Ss*, recognition scores for each word under each condition were obtained.

Procedure. The outline of the procedure is best conveyed by quoting the instructions read to *Ss*.

> The experiment is in three phases. First, I am going to [illegible] ability to recognize words that are presented to you in a very brief flash. Secondly I will give you a short list of words to [illegible]. I will test your recognition of [illegible] of words on brief exposure, some of the words being among those you have learned. Please place your face against the rubber visor on this apparatus, and view the white card. When I say 'now,' focus between the lines and when you're ready press the button once to flash up the word. Then, immediately say aloud and distinctly what you think the word was. Guess if you're not sure, and if you have no idea what it was, say 'no.' [practice phase]
>
> For the next phase, there are 12 words printed on this card in a random order. You have three minutes to learn them and go over them in your head. You will need to keep them in your mind through the third phase. Begin now. [pretraining]
>
> Time is up, and it is time for the third phase, but try to remember those words. Some of the following series of words will be among the words you have learned, some will be similar to those you have learned, and others will be quite different words. The procedure is the same as before. When I say 'now,' focus between the lines . . . . [testing]

In the practice period, about 100 word exposures were given, of varying durations, to familiarize *S* with the task. The shortest duration at which words were correctly recognized at the end of the practice was taken as the individual *S*'s baseline for testing.

The set of words for testing was presented in a random order at baseline flash duration, then in the same order again at 5 msec. longer duration, and so on for nine serial presentations. Baselines varied from 20 msec. to 60 msec., and each *S* received exposures increasing in duration to 40 msec. above his baseline.

All responses in the test were recorded, but no knowledge of results of any kind was given.

Results and Discussion

There were 39.8% correct responses, 48.1% error responses and 12.1% "no" responses. Only 7.0% of the error responses were not real words. Of these, nearly three-quarters were from four of the *Ss*, and five *Ss* gave no such responses. In general, error responses were similar in word length and gross shape to their eliciting test words.

TABLE 2

Recognition Performance

Scoring measure	Pretraining condition			
	S	M	V	O
Mean recognition score for words	86.8	63.3	51.1	46.8
Mean recognitions per 9 presentations	4.35	3.72	3.25	3.02
Percent recognition of root morphemes	49.9	46.4	36.8	36.6
Percent recognition of suffix morphemes	62.2	54.2	51.6	51.0

Note. See text for explanation of pretraining conditions.

It was decided not to use a "threshold" measure for comparison of scores, since it would depend heavily upon the arbitrary criterion set by *E* and would neglect much of the available data. Instead, a measure of performance was devised which gave a value to every correct response according to exposure difficulty, normalized across *Ss*.

First, each correct response was given a value equal to the reciprocal of the proportion of correct responses at that exposure over all *Ss*. This weighting gives higher values to more difficult discriminations. Thus weighted, correct responses were summed over the nine exposures to give a score for each word by each *S*. The scores of each *S* were further weighted by the multiplier necessary to correct his total score to 100. This procedure compensated for individual differences.

The mean recognition scores under the four conditions are shown in the first row of Table 2. The untreated data are summarized in the second row of Table 2 as the mean number of wholly correct recognitions per nine presentations under each condition. No further reference will be made to the latter figures.

A three-factor analysis of variance was carried out on the recognition scores

(model from Winer, 1971, p. 380). The factors were conditions, words, and replications (i.e., the two sets of eight *S*s duplicating each result). The pretraining conditions differentially affected word recognition, $F\ (3, 93) = 12.65$, $p < .001$.[3] The result of principal interest is that although the difference between scores under Conditions V and O fell well short of reliability ($F < 1$), Condition M did facilitate word recognition compared with O, $F\ (1, 93) = 5.38$, $p < .025$. However, the effect of Condition M was rather less than that of S, $F\ (1, 93) = 10.79$, $p < .01$. The Replications × Conditions interaction was not significant ($F < 1$), indicating that the differences between conditions were consistent between the two groups of *S*s. Similarly, the Words × Conditions interaction was not significant, $F\ (93, 93) = 1.32$, indicating that the conditions had the same effects over all the test words used.

There is some indication of effects of phonological similarity, which may be worthy of study in their own right. In six of the rows of words in Table 1, the "root" of Word 3 differed phonologically from those of Words 1 and 2, e.g., reader, reading, ready. A separate analysis of the recognition of the 12 test words from these rows indicated that under Condition V the phonological dissimilarity tended to counterfacilitate recognition. However, the analysis of variance of recognition scores was repeated with these 12 words excluded, and even when the pretraining and test word roots had the same phonology under Condition V, there was no significant facilitation compared with O, $F\ (1, 57) = 1.20$.

The results confirm the effectiveness of the pretraining technique for creating a short-term, experimentally controlled word-frequency effect. This effect does not occur when the pretraining word resembles the test word in visual–acoustic features only. There is reliable facilitation only when there is morphemic identity in addition to visual–acoustic similarity.

Considered in terms of the logogen model, this means that the logogen whose threshold is lowered by pretraining with a word corresponds not to that word, nor even to the visual–acoustic pattern it contains as root, but strictly to the morpheme from which the word is derived. Thus, the word recognition process, even in the absence of contextual or semantic stimulus information, involves assignment of the stimulus to a particular morpheme with its specific semantic associations. The morpheme, and only that one morpheme, to which the decision process assigns the word, has its logogen threshold lowered. Thus, when a pretraining word belongs to a different morpheme from a similar test word (as in Condition V), no facilitations would be expected. This interpretation would also account for the negative results of Ross et al. (1956).

The conclusion that the unit involved in facilitation effects is the morpheme was further illustrated by the percentage of responses containing the same root morpheme as the eliciting test word. These results are shown in the third row of Table 2. Bearing in mind that the recognition of whole words was considerably better under Condition S than under M, it is significant that the recognition of root morphemes under these two conditions did not differ reliably. It may be concluded that the latter result was due to the equal facilitation of root morphemes in these two pretraining conditions. Furthermore, recognition of root morphemes was not reliably different under Conditions V and O. This would be predicted, since under Condition V the root morpheme of the test word did not occur in pretraining and thus could not have been facilitated.

Even though the root *letter sequences* in pretraining were the same in Conditions M and V, with suffixes differing from the test words in both cases, the recognition of root *morphemes* was better under M ($t = 2.38$, $p < .025$). The facilitative effects of V pretraining were apparent in the intrusion

[3] In order to avoid the "language-as-fixed-effect fallacy" pointed out by Clark (1973), the appropriate quasi-F ratio for the conditions main effect was also calculated: $F'\ (3, 17) = 8.03$, $p < .01$. This provides statistical support for generalizing the effect to any comparable sets of words.

errors containing the root morpheme of the pretraining word. Such errors amounted to 8.7% of the responses under V, a figure comparable with the difference in root morpheme recognition between M and V. This is evidence that response biases of similar magnitude were created under these two conditions.

There appears to be some validity in considering experimental words in terms of root and suffix. The results suggest that, to some extent, free morphemes and suffixes are recognized independently. This is consistent with the linguistic analysis of words into morphemes, and the view that, the response units made available by logogens are morphemes.

It appears that the difference in word recognition scores between Conditions S and M was due to the pretraining with the test word suffixes under S, which did not occur in the other conditions. The scores for recognition of suffix morphemes are shown in Table 2. The difference in suffix recognition between S and O ($t = 2.56$, $p < .025$) demonstrates suffix facilitation. The differences in suffix recognition between M, V, and O were not reliable. The difference between S and M was only poorly reliable ($t = 1.59$, $p = .07$), probably because suffix recognition under M was inflated by the constraint imposed on response suffixes if the pretrained root was correctly produced. Thus, if the test word were "newer," then the likelihood of producing the correct suffix "-er" is zero if the root has been reported as "newt." If "new" has been seen correctly, the likelihood of guessing "-er" is appreciable.

Of the errors which were real words given in response to test words having the zero morpheme (i.e., no suffix), 213 (93.0%) had no suffix. These responses were not counted in the suffix recognition scores of Table 2, which are therefore an underestimate.

It is clear that there was substantially better recognition of suffixes than of roots. It may simply be that the suffixes were more distinctive visual patterns. The suffix "-ing" was correctly recognized on 81.5% of the trials in which it occurred in the test word, including 35.0% of the trials on which the first part of the response was wrong. For the suffix "-ed," there was recognition on 65.1% of its occurrences, including 28.3% error responses. Correct recognition of "-er" was 53.7%, including 15.3% errors.

It may be that "-ing" is more distinctive than "-ed." However, it is an attractive proposition that the suffixes themselves are subject to a frequency effect. If this were so, the relative proportions of trials on which the root and suffix of a test word were recognized would reflect the relative familiarity of the two morphemes to the subject.

This experiment gives grounds for considering a well-known phenomenon in new terms, as "the morpheme-frequency effect."

REFERENCES

Broadbent, D. E. Word-frequency effect and response bias. *Psychological Review*, 1967, **74**, 1–15.

Broadbent, D. E. *Decision and stress.* New York: Academic Press, 1971.

Catlin, J. On the word frequency effect. *Psychological Review*, 1969, **76**, 504–506.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–359.

Daston, P. Perception of idiosyncratically familiar words. *Perceptual and Motor Skills*, 1957, **7**, 3–6.

Howes, D. H. On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology*, 1954, **48**, 106–112.

Morton, J. A preliminary functional model for language behavior. *International Audiology*, 1964, **3**, 216–225. (Reprinted in R. C. Oldfield & J. C. Marshall, Eds., *Language*. London: Penguin Books, 1968).

Morton, J. A retest of the response bias explanation of the word frequency threshold effect. *British Journal of Mathematical and Statistical Psychology*, 1968, **21**, 21–33.

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, **76**, 165–178.

Nakatani, L. H. Comments on Broadbent's response bias model for stimulus recognition. *Psychological Review*, 1970, **77**, 574–576.

Neisser, U. An experimental distinction between perceptual process and verbal response. *Journal of Experimental Psychology*, 1954, **47**, 399–402.

Ross, S., Yarczower, M., & Williams, G. M. Recognitive thresholds for words as a function of set and similarity. *American Journal of Psychology*, 1956, 69, 82–86.

Rubenstein, H., & Pollack, I. Word predictability and intelligibility. *Journal of Verbal Learning and Verbal Behavior*, 1963, 2, 147–158.

Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition threshold for words. *Journal of Experimental Psychology*, 1952, 43, 195–201.

Treisman, M. On the wor[illegible] uency effect: Comments on the papers by [illegible] n and L. H. Nakatani. *Psychological* [illegible] 1971, 78, 420–425.

Tulving, E., Mandler, G., & [illegible] l, R. Interaction of two sources of info[illegible] in tachistoscopic word recognition. *C*[illegible] *Journal of Psychology*, 1964, 18, 62–71.

Winer, B. J. *Statistical princ*[illegible] *xperimental design.* (2nd ed.) New York: [illegible] w-Hill, 1971.

(Received July 12, 1

Journal of Experimental Psychology
1974, Vol. 103, No. 6, 969–974

CONSONANT–VOWEL–CONSONANT RECOGNITION AS A FUNCTION OF GRAPHIC FAMILIARITY AND MEANING [1]

SETH N. GREENBERG [2]

Ohio University-Lancaster

The present research determined whether perception of letter strings is influenced by meaning associated with the string and/or the familiarity of the string's graphic structure. Combining a familiarization procedure which allows graphic familiarity and meaning to be independently associated with unfamiliar consonant–vowel–consonants (e.g., Taylor) and a forced-choice word perception procedure (e.g., Reicher) which eliminates nonperceptual influences, it was demonstrated that only graphic familiarity improves letter sequence recognition. However, it was noted that meaning pretaining influenced the choice of a unitization strategy used in the perceptual task. The results indicate further information about the scope of variables influencing unitization strategy, as well as the importance of unitization strategy in the perception of certain fixed orthographic structures.

Using visual thresholds as measures of perceptibility, Cattell (1885) determined that man perceives letters more easily when the letters form words. A basic question that arises from Cattell's observation is which feature or features associated with words and not with random letter sequences contribute to the superior visual recognition of words? Gibson (1972) classified the distinguishing features of words into four types of information: graphological, phonological, semantic, and syntactic. The two features of interest in the present research include graphological–orthographic features which refer to the physical character and arrangement of letters in a sequence and semantic features. Specifically, this investigation is concerned with whether meaning and/or familiarity of a graphic letter sequence increases a sequence of letters' perceptibility.

Attempts to determine whether graphic familiarity and meaning influence letter sequence perceptibility face two major problems:

1. Employment of an experimental procedure that successfully eliminates nonperceptual factors from biasing results. For example, if the "perceptual task" requires *S*s to report a string of letters, after a brief visual presentation performance may reflect both what *S*s saw and what *S*s were able to recall correctly (Baddeley, 1961; Hochberg, 1968). Thus, a significant effect attributed to a feature might be due to that feature facilitating recall rather than perception. The problem of nonperceptual factors can be circumvented by the use of a forced-choice recognition technique introduced into word perception by Reicher (1969). Reicher's technique calls on *S*s to select from two response alternatives the letter that occupied a randomly designated position in the tachistoscopically presented letter sequence which immediately preceded the response choices. Both response alternatives are equally likely to appear in the tested position. Reicher's procedure successfully eliminates factors of storage load, interference during retrieval, and biased responding which hindered perceptual interpretations of previous experiments.

[1] This study is based on a dissertation submitted to the Ohio State University in partial fulfillment of the requirements for the PhD degree. The research was carried out at the Human Performance Center and was sponsored in part by National Institute of Mental Health Grant 1-R01-MH20495-1 to Harvey G. Shulman. The author wishes to thank Harvey G. Shulman for the time he devoted to reading earlier drafts and the many helpful ideas he contributed.

[2] Requests for reprints should be sent to Seth N. Greenberg, Department of Psychology, Ohio University-Lancaster, 1570 Granville Pike, Lancaster, Ohio 43130.

TABLE 1

PERCEPTUAL STIMULI GENERATED FROM THE FAMILIARIZATION CONSONANT-VOWEL CONSONANT ZEL

Beginning added vowels		End added vowels	
Intact	Nonintact	Intact	Nonintact
A ZEL	A VEL	ZEL A	VEL A
I ZEL		ZEL I	ZAL I
O ZEL	O ZAL	ZEL O	
U ZEL	U ZED	ZEL U	ZED U

Note. The actual set of perceptual stimuli used incl[illegible] randomly selected intact stimuli [illegible]

[illegible] Manipulation of one feature associated with a letter sequence is normally accompanied [illegible] another feature which is positively correlated with [illegible] in the primary feature. [illegible] better than MBH which is not [illegible] has been to [illegible] perception (Gibson, Bishop, Schiff, & S[illegible], 1964). However, the increased [illegible] of the letters I, B, and M when [illegible] to [illegible] i.e., graphic familiarity [illegible] simultaneous increase in [illegible]

[illegible] perception of [illegible] Taylor's results supported a [illegible]

iarity effect, but failed to [illegible] effect [illegible] meaning. However, use[illegible] threshold technique which [illegible] nonperceptual influences, [illegible] during recall and response [illegible] interpretation of Taylor['s] [res]ults. [In] addition, CVCs were pa[ired] with [illegible] definitions for only 15 trials [prior] to be[ing] employed in the perceptual [t]ask. [This] failure of the meaning varia[ble] [could] be [at]tributed to insufficient asso[illegible]

The present research foll[ow]ed a more extensive version of Taylor'[s] [famil]iarization procedure with a [illegible] recognition task in order [to] [de]termine whether graphic familiarity an[d] meaning are features that enhance a l[etter sequence's] perceptibility.

METHOD

Subjects. Twelve Ohio State Uni[versity] [illegible] served for 13 45-min. sessions at $1 [illegible] The Ss were assigned to three famili[arization] [illegible] (H = high meaning associate, L [illegible] associate, and P = no associate) in o[illegible] ance, thereby providing four Ss per g[roup].

Stimulus material and design. Tw[illegible] meaningful (m) CVCs (Archer, 1960) [illegible] stimuli in the familiarization task a[illegible]

[illegible]

Regarding all aspects of the perceptual [treat]ment, List 1 and List 2 were treated identically. The list used by a given S during familiarization would be the [illegible] stimuli, while [illegible]

[illegible] CVC [illegible] a set of 14 perceptual stimuli [used] in the experimental (visual recognition) task. [An] [illegible] of a perceptual stimulus set is exhibited [in Table] 1.

[The] first and third columns in Table 1 contain [illegible] which were constructed by [illegible] the four vowels that differs from the CVC's [illegible] beginning or end of the familiarization [illegible], each CVC generated eight intact stimuli; six intact [illegible] each set of eight [for] the purpose of testing the [illegible] positions. One stimulus from each set of six [illegible] was used during a session, providing 24 tests of [illegible] positions of intact stimuli. Counterbalancing [illegible] the 24 tests each session produced an equal n[illegible]

[illegible] of each CVC position. A different member [illegible] set of six stimuli was probed in each of six [illegible] sessions. Tests of the CVC positions were counterbalanced within each set so that each [illegible] was probed twice.

In order to meet the requirement of equally likely [illegible] alternatives for the forced-choice probes, [illegible] intact stimulus used for a CVC position probe had a nonintact partner which was derived by replacing the probed letter of the intact stimulus with another letter (second and fourth columns). The replacement letter in the nonintact stimulus and the letter replaced from the intact stimulus [illegible] alternatives for that particular [illegible] the intact stimulus and its [illegible] with an equal [illegible] position.

[illegible] each session, [illegible] of one [illegible] vowel and [illegible] vowel from [illegible] added vowel [illegible] All eight [illegible] that both vowels serving as response alternatives would be used as the correct alternative (i.e., occur as the added vowels with a given CVC equally often).

The rationale behind added vowel testing was that it permitted comparisons between familiar and unfamiliar stimulus conditions. The CVC [illegible] data could not be used for this purpose, since [illegible]

[illegible] perceptibility and upon scanning strategy (unitization).

The order of perceptual stimulus presentation and testing was random within each session (96 trials) except for the following constraints: (a) no successive repetition of stimuli containing the same CVC base was allowed, (b) no more than four successive tests of CVC positions or added vowel positions was allowed, (c) no more than four successive presentations of stimuli generated from CVCs belonging to one of the two lists was permitted, (d) over 32-trial blocks, half of the stimuli were List 1 and half were List 2 generated stimuli, and (e) no CVC appeared as the base of a perceptual stimulus more than twice in a 32-trial block.

[illegible] Ss in the H and L groups [illegible] 12 CVCs belonging [illegible]

[illegible] were used to establish a 90% [illegible] The method for determining [illegible] a short-cut psychophysical procedure [illegible] (1960). The test items [illegible] duration that [illegible] recognition threshold [illegible] was used during the [illegible] in the visual recognition task [illegible]

[illegible] preexperimental tasks, familiarization and [illegible] determination were performed [illegible]

TABLE 2

POSITIONAL ERROR RATES (IN PERCENTAGES)

Condition	Left-added vowel	Left consonant	Center vowel	Right consonant	Right-added vowel
Group P					
Familiar-intact	28	11	17	29	36
Familiar-nonintact		18	33	13	
Unfamiliar-intact	31	15	35	36	39
Unfamiliar-nonintact		16	26	11	
Σ	29.5	15	27.75	[illegible]	37.5
Group L					
Familiar-intact	17	19	15	18	46
Familiar-nonintact		24	32	39	
Unfamiliar-intact	22	17	28	29	49
Unfamiliar-nonintact		25	27	24	
Σ	19.5	21.25	25.5	27.5	47.5
Group H					
Familiar-intact	13	14	8	20	37
Familiar-nonintact		32	25	32	
Unfamiliar-intact	17	16	15	32	44
Unfamiliar-nonintact		8	22	17	
Σ	15	17.5	17.5	24.75	40.5

Note. Error rates listed under added vowel columns were collected over last four sessions only. Analyses [illegible] performed on added vowel data collected over last four sessions and all six sessions produced the same significant effect, [illegible] collected over last four sessions appeared more stable. Added vowel tests were completely counterbalanced within sessions.

forward association trial, while an *S* in the P group pronounced the list of CVCs three extra times. These familiarization trials preceded 18 final pronunciation trials (P group), or 18 forward association trials (L and H groups), as well as a final set of duration threshold trials.

The final six sessions were given to the experimental task, CVC plus vowel recognition. During a perceptual recognition session, an *S* sat 76 cm. in front of a 91 × 61 cm black screen with a 7.6 × 17.8 cm white Lucite area serving as the projection area. The projection area was long enough to hold seven projected letters with a .64-cm. space between the end position letter spaces and the black screen. The field was high enough to hold three rows of letters with a .32-cm. space between rows and a .42-cm. space between the top or bottom row and the edge of the black screen. The *S* pressed a button when he considered himself fixated at the center of the field. Immediately after the button press the stimulus appeared for a duration equivalent to *S*'s 90% recognition threshold for single letters plus 50 msec. The vowel position of the CVC portion of the stimulus coincided with the center of the fixation field. The added vowel appeared in either the leftmost or rightmost position. Thus, there was a space between the CVC base and the added vowel. Following termination of the stimulus display, a masking field was presented containing a series of overlapping Xs and Os, which covered the middle row where the stimulus had just appeared. Occurring with the masking field were two response alternatives which occupied positions directly above (top row) and below (bottom row) the position formerly filled with the stimulus' to-be-tested letter.

During the appearance of the response-alternative-masking field, *S* chose, from the two alternatives, the letter he thought occupied the probed position. Response alternatives remained available for 3 sec. The *S*'s choice was written on an answer sheet, which was covered following the response. The time between successive responses depended upon *S*'s response time, with a normal interval of about 15–20 sec.

The basic projection equipment included two Kodak Carousel RA950 projectors located behind blade shutters which were controlled by a Hunter 1516 six-channel timer.

Eleven of the 12 participants in the perceptual task were contacted 10 wk. after the completion of the perceptual experiment. The twelfth *S*, who had moved, was in the pronunciation group. Two tests were administered; a recall test first and then a recognition test. In the recall test *Ss* were asked to recall as many of the 12 familiar CVCs as they could, while in the recognition task *Ss* chose the 12 familiar CVCs from the total set of 24 CVCs. In both conditions, *Ss* were allowed as much time as they wanted to complete their responses. Each *S* was paid $1.25 for the recall–recognition session.

RESULTS

Error rates for all five tests positions are summarized in Table 2. In examining the influence of graphic familiarity on sequence perceptibility, only added vowel performance was considered. The added vowel results provide evidence that both left and

right [vo]wels were recognized more often when [di]splayed with a familiar CVC than with an unfamiliar CVC. An analysis of variance performed on the error data co[nfir]med this observation revealing a si[gn]ificant effect of graphic familiarity, [F (]1, 9) = 8.40, $p < .05$.

Had meaning influenced stimulus per[c]eptibility, *Ss* in the H and L groups should have performed significantly better than *Ss* in the P group. Examination of the CVC error data in Table 2 appears to indicate that meaning affected CVC per[c]eption. However, an analysis of variance [pe]rformed on CVC position data, as well [as a]n analysis of variance on added vowel [da]ta indicated that the between-*Ss* variable, [me]aning, failed to produce a significant [e]ffect, F (2, 9) < 1 for both analyses. The failure of the main effect of meaning could be due to inclusion of intact unfamiliar CVC tests and nonintact familiar and unfamiliar CVC tests with the intact familiar tests. Since only intact familiar CVC bases were treated differently in the various groups, i.e., used during familiarization, the effect of meaning should be confined to the intact familiar condition within each group, and would be identified as a significant Meaning × Familiarity × Intactness interaction. Analyses of [v]ariance performed on CVC position and [a]dded vowel position data failed to produce [th]is interaction, F (2, 9) < 1 for both [c]omparisons. The lack of an effect of meaning on the total perceptibility of a [se]quence could have also resulted from a [fa]ilure of the familiarization procedure to produce a difference in the CVCs for the various groups. Results from the long-term recall task place doubt on this possibility. The *Ss* in the H and L groups recalled an average of 11 and 9 of 12 CVCs, respectively, while those in the P group recalled 5 CVCs. Recognition performance did not differ betweeen groups. A set of tests performed on the recall results demonstrated that both association groups recalled more items than the P group ($p < .05$ for both comparisons). This information indicates that differential familiarization effectively established differences in the CVCs

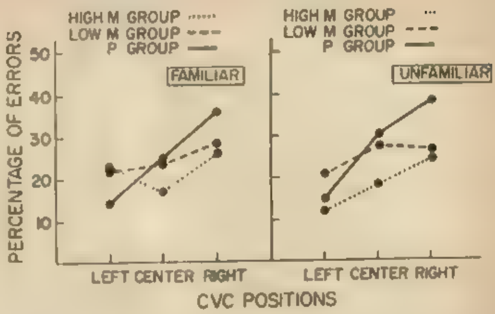


FIGURE 1. Meaning × Consonant–Vowel–Consonant (CVC) Position × Familiarity interaction. (Abbreviations: M = meaningful and P = no associate.)

for the various groups and thus eliminates the failure of familiarization as an explanation of the null effect of meaning.

While meaning failed to effect the level of CVC recognition, a significant Meaning × Position interaction, F (4, 18) = 14.59, $p < .05$, indicated an influence of meaning on the way CVCs were unitized. The interaction displayed in Figure 1 appears to result from the marked increase in errors from the left to center to right CVC position associated with the P group, contrasted with a considerably more moderate left to right increase in errors for the H and L groups. The slopes of the three curves in Figure 1 were compared with a slope of zero (see Draper & Smith, 1966, pp. 15–24). Only the slope describing the increasing errors across positions for the P group differed significantly from a zero slope, t (1) = 6.2, $p < .05$. In addition it should be noted that the positional error trends within each group were similar for familiar and unfamiliar CVC bases.

Discussion

The present results confirmed Taylor's (1958) finding of an effect of graphic familiarity upon letter sequence perceptibility. While meaning pretraining did not produce a similar effect, it did influence the strategies for unitizing the letter sequences. The *Ss* in the P group appeared to treat each letter of a CVC as a unit, and processed the letters in a left to right direction. In contrast, the requirement of making a meaningful association (H and L groups) encouraged *Ss* to treat the whole CVC

as a perceptual unit. Both strategies occurred with unfamiliar CVC-based stimuli, as well as with familiar stimuli. The effect of meaning pretraining, particularly with regard to unfamiliar perceptual stimuli, supported and broadened an earlier finding of Aderman and Smith (1971). Aderman and Smith demonstrated that pretraining involving the consistent presentation of letter strings with a certain orthographic structure could influence the unitization strategy used in processing a succeeding letter string. Two orthographic structures were used during pretraining with half of the *S*s receiving a list of stimuli having one structure, while remaining *S*s received a list with a second structure. In the present study, orthographic structure of both the pretraining stimuli and the base of the perceptual stimuli was held constant, i.e., CVC, while the level of meaningfulness was varied during pretraining. The fact that variations in meaning pretraining led to different unitization strategies is evidence that types of pretraining less closely associated with the physical structure of a letter sequence than orthographic variations can produce changes in strategies for unitizing strings of a fixed structure.

Aderman and Smith's (1971) results also indicated that if the pretraining orthographic structure conformed to the rules of English spelling patterns, *S* would attempt to unitize the succeeding test string into groups of letters identified as spelling units. If the succeeding string also matched an orthographic structure of English, the spelling unit strategy would be compatible with the succeeding string, and *S* performed better in this condition than if pretraining led *S* to use a strategy in which single letters served as units. Aderman and Smith found no advantage, however, to unitization strategies resulting from pretaining sequences either conforming or not conforming to English structure when the succeeding test string did not conform to English structure. Perception of CVCs also appears to be unaffected by unitization strategies resulting from pretraining. Whether CVCs were unitized as wholes or by single letters, the level of recognition was not significantly altered. Thus, various types of pretraining produce different unitization strategies, but whether those strategies influence the perceptibility of a letter sequence depends upon the orthographic structure of the test sequence. It should also be noted that some pretraining variables, such as familiarity in the present experiment, influence sequence perceptibility without affecting unitization (no interaction between familiarity and position was found). Hence, variations in unitization strategies are not necessary to produce changes in the level of perceptibility of letter strings of a fixed structure.

REFERENCES

ADERMAN, D., & SMITH, E. E. Expectancy as a determinant of functional units in perceptual recognition. *Cognitive Psychology*, 1971, **2**, 117–129.

ARCHER, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74** (10, Whole No. 497).

BADDELEY, A. D. Immediate memory and the "perception" of letter sequences. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 364–367.

CATTELL, J. M. Ueber der Zeit der Erkennung und Benennung von Schriftzeichen. Bildern und Fargen. *Philosophische Studien*, Wundt, 1885, **2**, 635–650. Translation in J. M. Cattell, *James McKeen Cattell, man of science*. Vol. 1. Lancaster, Pa.: Science Press, 1947

DRAPER, N. R., & SMITH, H. *Applied regression analysis*. New York: Wiley, 1966.

GIBSON, E. J. Perceptual learning and the theory of word perception. *Cognitive Psychology*, 1972, **2**, 351–368.

GIBSON, E. J., BISHOP, C. H., SCHIFF, W., & SMITH, J. Comparison of meaningfulness and pronunciability as grouping principles in the perception and retention of verbal material. *Journal of Experimental Psychology*, 1964, **67**, 173–182.

HOCHBERG, J. In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception*. New York: Holt, 1968.

REICHER, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, **81**, 275–280.

TAYLOR, J. A. Meaning, frequency, and visual duration threshold. *Journal of Experimental Psychology*, 1958, **55**, 329–334.

VANDERPLAS, J. M., SANDERSON, W. A., & VANDERPLAS, J. N. Statistical and associational characteristics of 1100 random shapes. *Perceptual and Motor Skills*, 1965, **25**, 414.

(Received July 14, 1973)

Journal of [illegible] Psychology
1974, Vol. [illegible] 975–980

[illegible]FORMATION EXTRACTION FROM DIFFERENT RETINAL LOCATIONS[1]

LESTER A. LEFTON
University of South Carolina

RALPH NORMAN HABER
University of Rochester

Four expe[illegible]ments examined manipulations of retinal location, visual field of presentation, [illegible] response variables in a visual matching task where the re[illegible] speed of judging whether the stimuli were the *same* or *dif*[illegible]. Results showed that (*a*) reaction time increased with retinal eccentricity, [illegible] ally no differences between the two visual fields, (*c*) a [illegible] retinal Location interaction emerged, and (*d*) both judg[illegible] sameness" and of "difference" were responded to in a similar manner, [illegible]rdless of the visual field of presentation.

[illegible] experiments examined an aspect [illegible] effects of retinal location on in[illegible]ation processing of tachistoscopically presented letters. In most of the literature, stimuli have been a horizontal row of letters spread across the fixation point, subtending generally less than 3° or 4° of visual angle. With the assumption that acuity would be sufficiently homogeneous over this region (the evidence summarized in Riggs, 1965, notwithstanding), retinal location has not been considered a particularly significant variable. This has been held even with the general findings in the same literature that accuracy of report, speed of reaction, clarity, apparent contrast, and other variables show variation over this range of locations. Eriksen, in a number of studies from his laboratory (e.g., Eriksen & Hoffman, 1972; Eriksen & Rohrbaugh, 1970) has worried about this sufficiently to avoid horizontal arrays altogether, preferring to place his stimuli in a circular arrangement around fixation. This is not always a feasible procedure, especially when recognition among groups of letters (e.g., words) is of interest.

Harcum (1964) has claimed that spatial organization among items in the array is a far more important variable than absolute retinal location. Hershenson (1969) has challenged this argument, suggesting that under circumstances in which the perceiver is thoroughly familiar with the organization of the material, the distance from fixation determines perceptual clarity and accuracy of report. Harcum (1970), after reanalyzing Hershenson's data, has not been convinced, and the issue is as yet unresolved.

In another line of work (e.g., Bryden, 1967; Heron, 1957), it has been shown that there is an asymmetry of accuracy of report across the fixation point. When a single letter appears either to the right or left of fixation, it is reported more accurately when on the right side, presumably because it projects to the left, language processing cerebral hemisphere. The recent findings of Egeth and Epstein (1972) complicate the issue. They presented a vertically arranged pair of capital letters either 4° to the right or 4° to the left of fixation. The *S*'s task was to respond with a *same* or *different* judgment by pressing a button which recorded re-

[1] The project presented herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Partial support for this research comes from the National Institute of Education (NE-G-00-3-0017) to the first author and from U.S. Public Health Service (MH 10753), Office of Education (OEG-0-72-0671), and from the Behavioral Research Directorate of the U. S. Army Human Engineering Laboratory (DAAD-05-71-C-0407) to the second author. We thank Nancy Belcher and Elaine Steele for help in data collection and John H. Wackwitz for help in statistical analyses.

[2] Requests for reprints should be sent to Lester A. Lefton, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208.

action time. When the letters were on the right, the *same* responses were faster than the *different* responses, and the converse was found for letters presented on the left. They attributed this difference to different processing occurring in the two hemispheres. Geffen, Bradshaw, and Nettleton (1972) required *S*s to make name matches as well as visual matches (Egeth and Epstein required only the latter). They found that name matches were faster when the letters were presented on the right visual field as might be expected from the language specialization hypothesis about brain function. However, they did not replicate the *same–different* interaction with visual field side.

This brief review is not intended to be exhaustive but only to show the range of variables relevant to problems of retinal location in information extraction tasks. There are many differences between the procedures of these and other related experiments, some of which are examined in the following studies. In each of the experiments reported, a pair of vertically arranged letters were briefly presented. The *S* had to classify them as rapidly as he could as *same* or *different*. The pair could appear in 1 of 11 locations varying up to 4° on either side of the fixation point. This arrangement permits an examination of the main effect of retinal location and the possible interactions between eccentricity and type of response (*same* or *different*) or visual field (left or right). In one of the experiments, an uppercase and a lowercase letter were always paired; *S* was required to make a name match rather than a physical match, thereby permitting assessment of this variable. Finally, one experiment compared the traditional reaction time procedures of using two response keys with that used by Egeth and Epstein (1972) in which *S* used only one button. There *S* was instructed to press a button for *same* and do nothing to indicate *different*, or vice versa.

GENERAL METH[illegible]

Subjects. In each of the experi[illegible]s, *S*s were undergraduate students at the Univ[illegible] of South Carolina who had normal or corre[illegible] vision and were right-handed. Experiment [illegible]d 4 students, Experiment IB used 60 st[illegible]; Experiments IIA and IIB used 34 an[illegible] students, respectively, and Experiments III [illegible] used 20 and 24 *S*s, respectively.

Stimuli and procedure. The stim[illegible] were pairs of letters presented directly abo[illegible] another. They were black Paratype uppercase [illegible] (11126). At a distance of 86 cm., they e[illegible]ubtended about .35° vertically and horizont[illegible] The vertical distance between the two le[illegible] was .15° (except in Experiment IIA; see below [illegible]e letters chosen were from zero-order appro[illegible]tion lists of letters. On half of the trials, th[illegible]uli were the same two letters; the remainin[illegible]als were two different letters. A background [illegible]nnel was always on except during the presen[illegible] of the letter pair. The luminance level in b[illegible] channels was equated at 2 ftL.

Upon introduction to the laborato[illegible] *S*s were informed about the nature of the tas[illegible] stimuli and were instructed to fixate on th[illegible]entrally located fixation spot. The stimuli [illegible] fixation were presented in a Gerbrands [illegible]stoscope (Model T-4B). Upon initiating a trial, [illegible]50-msec. stimulus would appear and a clock counter (Electronic Research Mfg., Model 2600) would start. The *S*s responded by pushing one of two buttons each 3.12 cm. from the start button and 3 cm. from each other. The *S*s used their index finger of the preferred hand for starting the tachistoscope and stopping the reaction time clock. The *S*s were counterbalanced across response buttons and each received one block of trials which were considered practice and not included in data analysis.

EXPERIMENT I

The main variables of study were the 11 different retinal locations and *same* or *different* judgment. These 22 experimental trials were arranged in blocks in a random order and repeated in a different order 10 times per *S*. The *S*s were first presented with one block of 22 trials which were considered practice, and then tested on the 220 experimental trials. Each *S* participated for seven sessions. The study was then replicated (Experiment IB) using many more *S*s with eight observations per *S* per condition, collected in only one session. In Experiment IA, all letters were used with random assignment to retinal locations. In IB, only the letters

O, A, N, F, I, C, T, R, and S were used, and they were counterbalanced over locations. In Experiment IA, the letters were placed at fixation or at .5°, 1°, 2°, 3°, or 4° either to the left or right of fixation. Experiment IB used the same locations except that the 4° location was replaced with 3.5°.

Results

The Ss made 7% and 5% errors in the two versions of the experiment, respectively; these trials were excluded from the reaction time analyses. Reaction times were scored and summed for each *S* across the conditions and entered into an analysis of variance, a different one for both versions of the experiment. In neither experiment was any main effect or interaction found with visual field side, nor were practice effects significant. The left side of Figure 1 plots these results; the right side collapses over visual field since that variable produced no effects or interactions in either version.

Judgments of *same* or *different* were equally fast—only a 2-msec. difference in both experiments. The different retinal positions, however, produced quite different reaction times, F (4, 12) = 29.2, and F (4, 236) = 36.7, respectively, both $p < .001$. In both experiments, retinal location and *same–different* response interacted, and in virtually the same way, as can be seen on the right-hand side of Figure 1, F (4, 12) = 3.3, $p = .03$, and F (4, 236) = 27.5, $p < .001$, respectively. Three of the reaction times near fixation for the *different* responses were slower than those of the *same* responses. In the eight other cases the *same* responses were slightly faster than the *different* responses.

Trend analysis applied to these data showed significant linear effects in both experiments for the position of the stimulus on the retina, F (1, 3) = 42.6, $p < .001$, and F (1, 59) = 101.6, $p < .001$, respectively. In Experiment IA, retinal location produced a quadratic component, F (1, 3) = 14.3, $p < .001$, and in Experiment IB, a cubic component was evidenced,

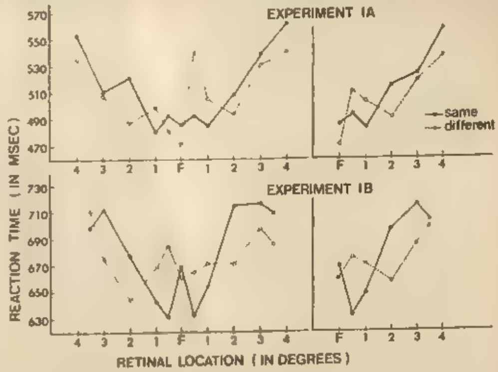


FIGURE 1. The mean reaction time (in milliseconds) is plotted as a function of the retinal location of the stimulus pairs in Experiment I. (The left panels show the data for all 11 locations, the right panels collapse the data over visual field of presentation.)

F (1, 59) = 8.2, $p < .006$. A critical trend is the interaction of the linear components of *same–different* judgments and retinal location—in both experiments these interactions were significant, F (1, 3) = 7.8, $p < .006$, and F (1, 59) = 27.3, $p < .001$. There were other significant and highly complex trends evident—in Experiment IB, the *same–different* linear component interacted with the retinal location quadratic component, F (1, 59) = 25.0, $p < .001$, and the left–right visual field linear component and the retinal location quadratic component also interacted, F (1, 59) = 10.5, $p < .001$. The triple interaction of the *same–different* and left–right linear components interacted with the cubic component of retinal location, F (1, 59) = 4.5, $p = .03$. These two interactions are highly complex, difficult to interpret, and account for a small portion of the variance (3.8%) relative to the linear trend of retinal location (51.8%) and the two-way interaction of *Same–Different* × Retinal Location (14%).

An analysis of the reaction times for erroneous responses was also undertaken. In general, error responses mirrored correct responses in most respects in that longer reaction times were more likely to be associated with errors. Thus, more errors were made for more peripheral locations. In the second experiment, more

errors were made for the *same* responses—8% vs. 2%, F (1, 59) = 62.3, $p < .001$. In both experiments, the frequency of erroneous reaction times produced an interaction between the *same–different* responses and retinal position, just as did the correct responses, F (4, 12) = 5.4, and F (4, 236) = 24.5, ps $< .001$ in both versions. There was a small interaction of Visual Field × Retinal Position in Experiment IB, F (4, 236) = 6.5, $p < .001$. The sum of squares for this interaction accounted for only 1.3% of the total variance.

Discussion

The results from both versions of experiments are consistent. Reaction time increased with peripheral eccentricity. Further, the pattern of errors generally reflected the reaction time data. Thus, where there were increases in reaction time there were corresponding increases in the number of errors. Since the error rate was generally low, *S*s were not making a speed–accuracy trade-off. When the letters were presented 3.5° or 4° in the periphery the processing task was more difficult, *S*s therefore made more errors and, in addition, took longer to process the items.

There was no evidence of an interaction of Visual Field × *Same–Different* responses, the results reported by Egeth and Epstein (1972). However, in Experiment I the letter pairs were separated only by a minimal amount of space. In the study reported by Egeth and Epstein, the letters were separated vertically by 4°. Experiment II was therefore constructed to examine this difference by placing pairs of letters either close together or far apart in either the left or right visual field.

Experiment II

Two vertically arranged letters were presented 3.5° either to the left or right of fixation. They were chosen from the symmetrical letters A, H, I, O, T, U, V, X, and Y. In one condition, the letters were separated vertically by 4°, similar to that procedure used by Egeth and Epstein (1972). In the other, only .15° separated the two letters. In both conditions there were two experimental conditions each with two levels: visual field and the response *same* or *different*. Since there were nine stimulus letters used, there were nine possible *same* combinations. Nine *different* combinations were then arranged and these nine were repeated in the reverse order, that is, with whatever was the top letter being the bottom letter. There were two visual fields, nine different stimuli, two possible responses (*same* compared with *different*), and four replications of each of the possible combinations making a total of 144 experimental conditions. These were arranged in blocks of 36 trials, each block being one replication of the 36 experimental trials. These four blocks were each arranged in a random order. At the end of an experimental session each *S* would have provided 36 measurements for each of the four possible conditions.

Results

There was a great deal of variability with respect to the number of errors made. Some *S*s made many errors, while others made very few. The *S*s were therefore divided into two groups—those *S*s who made 10% errors or less and those who made more than 10% errors. In data analysis, these two groups have been treated separately. However, except in the one instance to be noted, the pattern of results was so similar across error rates, that *S*s have been combined with respect to the presentation of results.

In general, no significant effects or interactions were found with either the wide (3.5°) or close (.15°) vertical separations. Specifically, no hint of a Visual Field × *Same–Different* interaction was apparent in either version, regardless of the error rate. These data are presented in Table 1. There were significantly more errors made on the *same* response for both error groups for both separations.

The results of this experiment confirm those of the first experiment. In this study, when the letters were presented close together so that *S*s could make a physical match, or far apart such that a name match might be made, there was no Visual Field × *Same–Different* interaction.

There [...] been no clear evidence in Experim[...] I, II, or in the Egeth and Epstein (1972) [...]udy that Ss were indeed naming the s[...]uli. Experiment III examined the nam[...] process by using capital and small lett[...] as paired stimuli.

Experiment III

The stimuli for this experiment (A, B, C, D, E, G, T, N, F, P, R, H, V) were [...] different set of letters chosen to reduce [...]he visual similarity between the upper- [...]d lowercases. The *S* was now to indicate [...]he letters had the same name or not. [...]ll other respects this study was iden[...]l to that of Experiment II. The [...]rtical separation was .15°.

Results

All *S*s tested in this experiment produced high error rates, averaging 27%. As in the other studies, the error rate was significantly greater for the *same* responses than for the *different* responses, F (1, 19) = 60.5, $p < .001$.

The reaction time data were treated as in the other experiments. There was a [...]ignificant left visual field superiority, [...] (1, 19) = 5.8, $p < .05$. Overall, the [...]t visual field was responded to 29 msec. [...]ster than the right visual field. There [...]as no difference between the *same* and *d[...]erent* response, nor any Visual Field [...] *Same–Different* interaction. These data are presented in Table 1.

[...]he critical interaction of Egeth and Epstein (1972) remains elusive. When *S*s were presented with stimuli that were named (Experiment III) or not named (Experiments I and II), or when stimuli were close together or far apart (Experiment II), this interaction did not emerge. There is one further methodological point that may be relevant. Egeth and Epstein used a go–no-go technique, in which half of their *S*s were asked to respond by pressing a single button as fast as they could if the letters were the *same* and do nothing if the letters were *different*. The other *S*s pressed if the two letters were *different* and did nothing if they were the *same*.

TABLE 1

Reaction Time (in msec.)

Experiment and condition	Visual field: Left	Visual field: Right
Experiment IIA		
Same	792	788
Different	774	768
Experiment IIB		
Same	783	801
Different	769	786
Experiment III		
Same	988	1,026
Different	967	987
Experiment IV		
Same	567	574
Different	603	600

Experiment IV

Experiment IV is a replication of Experiment II, differing only in that half of the *S*s were told to press the *same* button if the letters were the same and do nothing if they were different. The other half of the *S*s were told to press the *different* button if they were different and do nothing if the letters were the same. The letters were separated by .15°.

Results

The data were treated in the same manner as in the other experiments. Overall, there was a 10% error rate. The analysis of variance showed that there was neither a main effect of *same* compared with *different* nor of the visual field of presentation. The Visual Field × *Same Different* interaction was not evident ($F < 1$). The error rates showed no significant differences or interactions. Thus, except for the overall reduction in reaction time between this experiment and the two-button ones, this methodological change cannot be responsible for the overall pattern of results in these experiments.

General Discussion

One of the main findings of the present experiments was the change in reaction time

that was associated with changes in the retinal location of the stimuli. Experiment I showed that with increases in retinal eccentricity from fixation, judgments of sameness and of difference required more processing time. Both Harcum (1964) and Hershenson (1969) would predict this to be the case—both accuracy and reaction time data should reflect one another. These data do not directly converge on the notion of sensitivity compared with saliency as put forth by Harcum (1970). They do provide data that can be used to estimate the effects of manipulating fixation in other reaction time tasks that have spread letters across a fixation point. For example, a certain number of milliseconds might be subtracted from different retinal locations to examine visual processing uncontaminated by varying retinal locations.

The data provide evidence for a change in processing strategy depending on the locus of stimulation. The *Same–Different* × Retinal Position interaction found in Experiment I shows that at near foveal positions *same* responses are faster than *different* responses; this is a result that has been found many times (e.g., Posner, 1969). As the retinal eccentricity increases, however, the *different* responses come to be faster than the *same* responses. These results, the obtained interactions, and trend analyses support the conclusion that *S*'s strategies may shift to a criterion of sameness or difference as a function of the location of stimulation.

None of the other variables examined by the experiments seemed to be critical. The previously reported interaction of visual field with *same–different* responses was never observed. The only visual field effect found occurred when *S*s had to make a name match, and this was equally strong for both responses. The amount of the vertical separation of the letters did not seem important, nor whether *S*s used two response buttons, or only one with a go–no-go criterion.

Overall the results of the present studies present data to support five main conclusions: (a) When *S*s are engaged in a visual matching task, performance is slower the further the stimuli are from fixation; (*b*) there is a change in the speed of responses for [illegible] responses compared with *different* respon[illegible]e to the fovea; (*c*) there is no differ[illegible] between judgments of sameness or di[illegible]ce with respect to visual field of present[illegible]; (*d*) the only hint of a visual field effect [illegible]ars when *S*s are forced to name the stim[illegible] and (*e*) neither spacing arrangements [illegible] response key arrangements seem to ex[illegible] critical influence on the results.

REFERENCES

BRYDEN, M. P. A model for the s[illegible]al organization of behaviour. *Canadia[illegible]rnal of Psychology*, 1967, **21**, 37–56.

EGETH, H., & EPSTEIN, J. Differen[illegible]ecialization of the cerebral hemispheres [illegible]he perception of sameness and difference [illegible]*ption & Psychophysics*, 1972, **12**, 218–220.

ERIKSEN, C. W., & HOFFMAN, J. E. [illegible]haracteristics of selective attention in visu[illegible]eption determined by vocal reaction time. [illegible]*ption & Psychophysics*, 1972, **11**, 169–171.

ERIKSEN, C. W., & ROHRBAUGH, J [illegible] Some factors determining efficiency of selec[illegible] attention. *American Journal of Psycholo[illegible]*[illegible]70, **83**, 330–343.

GEFFEN, G., Bradshaw, J. L., & NETTL[illegible], N. C. Hemispheric asymmetry: Verbal a[illegible] spatial encoding of visual stimuli. *Journal [illegible] Experimental Psychology*, 1972, **95**, 25–31.

HARCUM, E. R. Interactive effects wi[illegible] visual patterns on the discriminability of individual elements. *Journal of Experimental Psychology*, 1964, **68**, 351–356.

HARCUM, E. R. Perceptibility gradients for tachistoscopic patterns: Sensitivity or saliency. *Psychological Review*, 1970, **77**, 332–337.

HERON, W. Perception as a function of retinal locus and attention. *American Journal of Psychology*, 1957, **70**, 38–48.

HERSHENSON, M. Perception of letter arrays as a function of absolute retinal locus. *Journal of Experimental Psychology*, 1969, **80**, 201–202.

POSNER, M. I. Abstraction and the process of recognition. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 3. New York: Academic Press, 1969.

RIGGS, L. Visual acuity. In C. H. Graham (Ed.), *Vision and visual perception*. New York: Wiley, 1965.

(Received July 23, 1973)

Journal of E… …tal Psychology
1974, Vol. … , 981–983

LINGUISTIC INTERDEPENDENCE OF BILINGUALS [1]

MIKE LÓPEZ [2] AND ROBERT K. YOUNG

University of Texas at Austin

Spanish English bilinguals were first familiarized on a list composed of Spanish or English words. Next they learned a Spanish or an English list consisting of words that were the same as (or translated) or different from those in the familiarization list. Familiarization effects were uniform both between and within languages in that the amount of positive transfer obtained was the same for all groups. The data were interpreted as being supportive of the language interdependence hypothesis.

The unique ability of bilinguals to process information presented in two different linguistic systems has given rise to differing ideas as to the nature of bilingual organization of memory. One of these, the independence hypothesis, assumes that the bilingual has two distinct memory stores, one for each language, with information presented in one language not being readily available to the other system. A contrasting position, the interdependence hypothesis, holds that bilinguals store information in one memory store only (Kolers, 1966). An interdependence hypothesis would assume that bilinguals store words in terms of their semantic features only, with some means of "tagging" the items with the proper language at the time of output.

Although several studies of bilingualism have supported the interdependence hypothesis, support for this position has not been unanimous. Most studies supporting the interdependence viewpoint have demonstrated a bilingual equivalence effect, wherein the presentation of an item in one language produces the same effect on recall of that item in the other language as if it had been originally presented in the other language (e.g., Glanzer & Duarte, 1971; Kolers, 1966). Tulving and Colotla (1970), supporting the independence hypothesis, maintain that this equivalence effect is probably due to a well-established habit of translating. Facilitation in the recall of translated items would not be evident were it not for an automatic translation of these items by *S* at the time of presentation. Presumably, this translation habit would be strongest in bilinguals who are not equally proficient in each language, translating from the weaker to the dominant language (Goggin & Wickens, 1971).

The present study was designed to test this translation hypothesis by giving *S*s familiarization trials on items in one language and then having them learn, by free recall, the same items in translation. If the bilingual's two languages represent two independent systems, there should be no familiarization effect at all, unless *S*s actively translate the items. However, this translation should result in facilitation of learning of items in the stronger language only, after they have been familiarized in the weaker language. There should be no facilitation effect for items familiarized in the strong language and learned in the weak language, since *S*s would have no reason to translate in this direction during familiarization.

The interdependence hypothesis would predict facilitation of learning in all cases. If the familiarization of a word or its translation results in the activation of the same semantic representation, then this familiarization should result in an equal facilitation of learning, regardless of whether or not learning is in the same language as familiarization, and regardless of which language is dominant.

[1] During the preparation of this article, the first author was supported by a Ford Foundation advanced-study fellowship, and the research was supported by Grant MH-19288 from the National Institute of Mental Health.

[2] Requests for reprints should be sent to Mike López, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.

TABLE 1

Mean Numbers of Items Recalled Per Trial

List language	Familiarization language: English	Familiarization language: Spanish
Experimental		
English	12.38	11.91
Spanish	10.25	11.39
Control		
English	11.15	11.35
Spanish	9.16	8.79

Method

Subjects. The *Ss* were 64 Mexican-American high school juniors and seniors, bilingual in English and Spanish, who volunteered to serve in the experiment. These *Ss* were asked to rate themselves on a number of variables related to language learning and fluency. Assignment of *Ss* to conditions was by a predetermined random sequence according to order of appearance in the experimental room.

Design. The *Ss* first read a 16-word list in English or Spanish and then learned a 16-word list in the same or the other language. In the experimental conditions, the second list consisted of the same words that were in the familiarized list, in the same language or translated to the other language. In the control conditions, the first-list words were unrelated to the words to be learned. Two *Es* tested the *Ss*. The design can thus be summarized as a 2 × 2 × 2 × 2 factorial design with a within-*Ss* variable of 10 free-recall trials (n = 4/cell).

Materials. The verbal materials used in the experiment were selected by asking 51 students from the same school as the experimental *Ss* to translate 100 common adjectives from English to Spanish or vice versa. Two different forms were used to counterbalance for direction of translation of the adjectives. The 32 adjectives most often translated correctly composed the four lists used in the study. None of the students asked to provide these translations served in this experiment.

Procedure. Each of the four 16-item lists was typed on a plain sheet of typing paper. As each *S* entered the experimental room, he was told by *E* that before the experiment started, *E* would like to hear *S* read a list of words. The *Ss* were then required to read the list aloud for six repetitions. Following the sixth repetition, *S* was told that the experiment could now begin and was read standard free-recall instructions. All *Ss* were then given 10 trials on the list. Stimuli were presented auditorily on tape at a 2-sec. presentation rate. Recall was oral with a 90-sec. recall interval. Order of presentation of the words was randomized across the 10 trials, and none of these orders corresponded to the order of the or al list read by *Ss* in the experimental conditio

Results

The *Ss* rated their fluenc each language on a 1–7 scale. The n English rating was 4.97, and the m rating of Spanish fluency was 3.77. addition, *Ss* indicated that they had sp English and Spanish for 13.27 and 53 years, respectively. Fifty-eight of the 64 *Ss* indicated that they thought English, χ^2 (1) = 34.51, $p < .01$. Th s appear to be compound bilinguals (c Ervin & Osgood, 1965) with English as ir dominant language. That is, these learned both languages at the same ti and in the same context, yet (probal due to a deemphasis of Spanish and a co esponding emphasis on the use of E glish in school) they presently consider th mselves to be more fluent in English.

An analysis of variance was performed on the number of items recalled o er trials. The experimental *Ss* recalled sig ficantly more items than did the controls, (1, 48) = 10.20. English lists were easier to learn than Spanish lists, F (1, 48) = 1 .3. In addition, the trials main effect ached significance, F (9, 432) = 124 All these *Fs* were significant at the .01 level. The Learning Language × Trials interaction was significant, F (9, 432) = 2.54, $p < .05$, as English lists were learned faster than Spanish lists. This is not surprising in view of the fact that for this group of bilinguals English is highly dominant.

A facilitation effect of familiarization was found for all language conditions, and this enabled experimental *Ss* to recall significantly more words than control *Ss*, as previously noted. Inspection of Table 1 reveals that there was not the expected differential facilitation that would have been predicted by a translation hypothesis. That is, familiarization of a Spanish list followed by learning the English equivalents does not result in a greater facilitation, relative to control items, than does familiarization of an English list followed by learning of the Spanish equivalents.

This differential facilitation would have been manifested in significant Familiarization Language × Learning Language and Familiarization Language × Learning Language × Condition interactions. These interactions were not significant, $Fs\ (1, 48) = .36$ and 1.62, respectively.

DISCUSSION

The results of the present study support an interdependence hypothesis of bilingual organization of memory. A positive transfer effect of familiarization was produced in all combinations of familiarization list language and learning list language, both intralingual and interlingual. The positive transfer in the interlingual combinations is especially revealing in view of the fact that there was no significant Familiarization Language × Learning Language interaction. That is, these English-dominant bilinguals did not show a significantly greater facilitation in the Spanish to English conditions vs. the English to Spanish conditions, as they should have if a translation hypothesis of the bilingual equivalence effect is to be accepted.

The greater difficulty *Ss* have in learning Spanish lists relative to English lists can be attributed in part to the learning history of these *Ss*. In school they are expected to learn in English, not Spanish. Consequently, they have not developed the elaborate coding strategies in Spanish that are necessary for rapid acquisition of verbal material.

Finally, it should be noted that a single semantic store for bilingual memory that is supported by this experiment does not exclude the possibility that other components, notably the syntactic and phonological components, of a bilingual's linguistic systems are independent. This aspect of bilingualism has been closely investigated by Macnamara (1967).

REFERENCES

ERVIN, S. M., & OSGOOD, C. E. Second language learning and bilingualism. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics: A survey of theory and research problems.* Bloomington: Indiana University Press, 1965.

GLANZER, M., & DUARTE, A. Repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 625–630.

GOGGIN, J., & WICKENS, D. D. Proactive interference and language change in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 453–458.

KOLERS, P. A. Interlingual facilitation of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 314–319.

MACNAMARA, J. The linguistic independence of bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 729–736.

TULVING, E., & COLOTLA, V. A. Free recall of trilingual lists. *Cognitive Psychology*, 1970, **1**, 86–98.

(Received August 2, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 984–991

ROLE OF HIGHER ORDER RULES IN PROBLEM SOLVING[1]

JOSEPH M. SCANDURA[2]

University of Pennsylvania

Why do some problem solvers succeed on problems for which they have all the necessary component skills, whereas others do not? It was assumed that a major portion of problem-solving ability can be traced to the presence or absence of higher order rules which make it possible to generate solution procedures. Experiments I and II were designed to test a simple problem-solving mechanism by which lower and higher order rules interact. Experiment I involved a composition higher order rule and Experiment II, a higher order generalization rule. Results were over 98% consistent with theoretical predictions concerning the behavior of individual *Ss* on specific problems. In Experiments III and IV the presence or absence of higher order rules was determined indirectly through testing. The overall percent correct predictions were 85 and 92 respectively.

One of the most crucial questions in problem-solving research is why some problem solvers succeed on problems for which they have all of the necessary component skills, whereas others do not. In dealing with this question, most studies of problem solving have attempted to deal with the process in its full complexity (e.g., Duncker, 1945; Newell & Simon, 1972). That is, most studies have employed problem situations that involve problem definition (the formation of subgoals), memory, the derivation of solution procedures, and the use of such procedures. The situation is further complicated because existing theories of problem solving are either extremely limited in scope (i.e., to specific kinds of problems) or, if general, are more like overall schemas than strong theories (cf. Greeno, 1973).

The present research has adopted a somewhat different strategy. It seeks to deal separately with the various aspects of problem solving (e.g., the derivation of solution procedures) but in a way which generalizes over tasks. Specifically, this research is not concerned with the role of memory or with problem definition. Precautions were taken to insure that *Ss* fully understood the problems (goals and givens) presented and that they were understood in the way *E* intended. Further, the ongoing research is concerned with a specific falsifiable theory (Scandura, 1973) concerning the behavior of individual *Ss* in particular problem situations, and not with either the performance of groups or of individuals averaged over tasks.

The general point of view adopted is that rules underlie all behavior (e.g., Scandura, 1970). More specifically, this research is based on the assumption that, when the goal adopted by *S* is known, a major portion of problem-solving ability can be traced to the presence or absence of higher order (HO) rules. These rules may be used to combine constituent parts of a problem solution into a coherent whole adequate for solving the problem (cf. Saugstad, 1955) and/or to otherwise generate solution procedures from known rules.

In this context, the question arises as to how known rules interact in problem solving. It is hypothesized that in problem solving (as in all learning) rules interact according to a fixed mechanism, which is presumed to be innate rather than learned (Scandura, 1973). This mechanism is

[1] This research was supported in part by National Science Foundation Grant GW-6796 and U.S. Office of Education Grant OEG-3-71-0136. Experiment I was conducted with Louis Ackler, Experiment II with Francine Endicott and John Durnin, and Experiments III and IV with Durnin and Wallace Wulfeck. The author would also like to thank Joe Karabinos and Diane Triman, who helped proctor the experimental sessions.

[2] Requests for reprints should be sent to Joseph M. Scandura, University of Pennsylvania, 3700 Walnut Street, Philadelphia, Pennsylvania 19174.

assumed to operate as follows: Given a goal situation ⟨S, G⟩, *S* tests to see if the problem solution is immediately available (i.e., to see if he knows the solution). If not, control shifts to the solution goal (SG), consisting of the set of potential solution rules, and *S* tests the rules available to him to see if any includes S in its domain and the goal in its range.[3] If such a rule is available, control reverts to the original goal and the rule is applied to S. If not, control shifts to the higher level goal (HG), [illegible] of HO rules which apply in the [illegible]n goal situation ⟨S, G⟩ and whose [illegible] contain a potential solution rule. With this higher level goal in force, *S* is assumed to test his available rules as before. If one is found, control reverts to the next lower level goal and the HO rule is applied to other available rules, thereby generating a new (potential) solution rule. If the lower level goal is satisfied by the new solution rule, control reverts to the original goal and the solution rule is applied to S. The problem is solved if the (potential) solution so generated satisfies the goal situation.

It should be noted that testing, goal shifting, and rule application are essentially the same at *all* goal levels. Generalization and formalization of the mechanism is given in Scandura (1973). Furthermore, HO rules are formally identical to other rules and obey the same laws of behavior. The descriptor "higher order" refers to the role of a rule in a particular instance of problem solving, not to its basic nature. This differs from common usage. The term "higher order rule" as used by Gagné (1970), for example, corresponds to the *output* of what is referred to here by this term.

Suppose, for example, that *S* adopts goal G in situation S and has only the following rules available: r_1, r_2 (neither of which satisfy SG), and h, where h (r_1, r_2) = p and p satisfies SG. What can we say about what *S* will or will not do? Given that *S* has adopted goal G in situation S, and only has rules r_1, r_2, and h available, Table 1 shows that the above mechanism provides a sufficient basis for predicting *S*'s behavior. If we changed the situation by eliminating any one of r_1, r_2, or h, we could conclude that *S* would fail. In order to make predictions where training on rules is not a central variable, it is essential to determine what rules *S* knows. Indeed, the general importance of prior knowledge in research on complex human behavior is becoming increasingly clear (e.g., Bourne, Ekstrand, & Dominowski, 1971; Scandura, 1973).

The major purpose of this research was to test this mechanism empirically, both by manipulating HO rules and by testing to determine their availability prior to problem solving.

TABLE 1

SEQUENCE OF EVENTS IN IDEALIZED PROBLEM SOLVING

Event	Theoretical justification
SG = $\{r_i \mid S \epsilon \text{ Domain } r_i, G \subset \text{Range } r_i\} = \phi$ (is empty)	Assumption
∴ Control shifts to HG	Higher level hypothesis
HG contains h	Assumption
∴ Control reverts back to SG and *S* applies h to r_1, r_2 (to get p)	Reversion and performance hypotheses
p satisfies SG	Assumption
∴ Control reverts to G and *S* applies p to S (to get R) R may satisfy G	Reversion and performance hypotheses

Note. See text for explanation of abbreviations used.

[3] This condition is slightly different from that in an earlier formulation (Scandura, 1973). Instead of requiring that the range of the rule contain the goal, I had initially proposed that the range be contained in the goal. While the latter condition works equally as well with initial goals, it is not fully adequate. For example, in the trading game experiment requiring the range of a composition rule to be contained in the higher level goal is overly restrictive because this goal only contains A → Y → C rules (Y variable and A and C constants), whereas the range of the composition rule more naturally includes all composite rules of the form X → Y → Z (where X and Z are also variables).

Experiment I

Experiment I provided a test of the mechanism involving an HO composition rule.

Method

Materials and tasks. The experimental material consisted of tasks involving trading stimulus objects of one kind (e.g., red chips) for response objects of another (e.g., pencils). Each task can be characterized as a set of stimulus-response pairs in which n stimulus objects are mapped into $(n + m)$ response objects (e.g., n red chips into $n + 3$ pencils). Throughout the experiment, $n + m \leq 10$, with $n \leq 7$ and $m \leq 4$.

There were two kinds of rules, *simple* and *composite*, for solving such tasks. Simple rules were represented on 5×8 in. cards. To represent the simple rule which maps n paper clips into $n + 1$ blue chips, for example, a paper clip was glued on the left and a blue chip on the right of the card. Composite rules generate trades in two steps. One composite rule, for example, first changes pencils into paper clips and then paper clips into white chips. Composite rules were represented by taping together two simple rule cards.

A pair of simple rules was said to be *compatible* if the outputs of one of the rules were of the same type as the inputs of the other. Compatible rules can be combined to form composite rules (e.g., the rules denoted "n caramels $\rightarrow n + 1$ toy soldiers" and "n toy soldiers $\rightarrow n + 2$ pencils" can be combined to form a composite rule which maps n caramels into $n + 3$ pencils). The set of compatible pairs of simple rules comprises the domain of an HO rule which maps such pairs into corresponding composite rules. This HO rule was used to define a second, higher level task in which, given a compatible pair (e.g., "n caramels $\rightarrow n + 2$ white chips," "n white chips $\rightarrow n + 1$ pencils"), the goal was to devise the corresponding composite rule (i.e., for converting caramels into pencils).

Subjects, design, and procedures. The *S*s were 31 boys and girls between the ages of 5 and 9. They were run individually and given 25¢ for participating. Seven *S*s were excluded from the experimental comparison, 6 because they passed the pretest and 1 because she failed to learn how to interpret the composite rule cards. At the start of the experiment, each *S* was told that he was going to play a trading game with *E* and was given several sets of objects to trade. The *S* was then taught to interpret simple rule cards and to make trades using the rules represented by the cards. For example, *S* was shown a rule card for trading paper clips for blue chips and *E* explained "no matter how many paper clips I give you, you must give me the same number of blue chips and then add one more." Practice was provided and, if necessary, *E* showed *S* how to make trades of this sort and asked him to repeat what he had been shown. The criterion was three consecutive successful trades. The *E* then gave *S* a set of objects not in the domain of the rule (e.g., two pencils) and asked *S* if he could use the rule to trade pencils for blue chips. Regardless of *S*'s response, *E* emphasized that the rule could be used only to trade paper clips for blue chips.

The *E* then showed *S* a different rule card and asked him to interpret the rule, providing assistance if necessary. Practice continued until *S* reached criterion. Again, *E* emphasized that the rule was restricted to objects on the card. This procedure was repeated using different rules until *S* was able to interpret three consecutive rule cards and use them without assistance. At this point, it was assumed that for *S* simply seeing a simple rule card was equivalent to knowing and being able to apply the rule.

During Phase 2, *S* was taught to interpret and use the composite rules. For example, *S* was presented with a composite rule for trading pencils for white chips together with two pencils. He was told to use the first part (*E* pointed at the first card) to trade the pencils for paper clips. The *S* was required to place the correct number of paper clips on the table. The *E* continued, "Now, use the second part to trade these four paper clips for white chips." This procedure was repeated with different stimuli until *S* performed three consecutive trades correctly. Practice continued with similar rules until *S* interpreted and correctly applied three consecutive composite rules.

The *S* was then given a *pretest* consisting of two parts. First, *S* was presented with new cards representing a pair of compatible rules. Then, *S* was asked to make three trades requiring use of the corresponding *composite* rule. (The *S* was never shown this rule directly, either before or after testing.) For example, an *S* who was presented with the rules "n pencils $\rightarrow n + 2$ pieces of bubble gum" and "n pieces of bubble gum $\rightarrow n + 1$ paper clips" would be presented in turn with various numbers of pencils (e.g., two, four, and one) and asked to trade the appropriate numbers of paper clips. No reinforcement was provided on the pretest.

Those *S*s who consistently failed on the pretest were randomly assigned, in matched pairs, to one of two treatment groups, Group HR, which received training on the HO rule, and Group C, which was the control group. Those *S*s who succeeded on two or more instances of the pretest were given 5 min. of irrelevant instruction (reading a comic book) and were given a posttest, which involved new rules but paralleled the pretest in every other respect.

Each *S* in Group HR was taught the HO rule. The *S*s were first shown two compatible rule cards and asked to interpret each. The *E* then demonstrated how to combine the rules by sliding the rule cards together in the appropriate manner. The *S* was then asked to interpret this newly formed rule, and *E* emphasized that the rules could be combined only because the output of one was the same as the input of the other. Next, *S* was presented with several pairs of rules, some of which were not com-

patible. [illegible] each pair, *S* was required to form the compo[illegible] rule, if possible, and to interpret the newly [illegible] rule but *S* did not actually make trades. [illegible] the randomly interspersed incompatible pairs [illegible] had to indicate that the rules could not be so co[illegible]ned. After performing successfully on five co[illegible]ive (compatible and incompatible) pairs of rul[illegible] *S* was given the posttest. The time required [illegible] *S* on the HO rule was recorded. Each G[illegible] C *S* was asked to read a comic book for the [illegible] amount of time as his matched partner in [illegible]roup HR required to learn the HO rule. Finally, the posttest was administered in the same manner as the pretest to all *Ss*.

[illegible]esults

[illegible] experimental results were relatively [illegible]. Eleven of the 12 HR *Ss* solved [illegible] ree transfer problems. The one *S* who [illegible] after reaching criterion on the HO [illegible] was put through the experimental [illegible]ocedure 1 wk. later. This time he performed perfectly on the posttest (only). All 12 control *Ss* failed uniformly on the posttest. The individual results are summarized in Table 2.

Although statistical comparisons between groups seem inappropriate, reliability of the percent correct predictions may be expressed in terms of confidence intervals (for the HR group). Based on the assumption that correct and incorrect predictions [illegible]re binominally distributed and using the [illegible]tained mean of 91.7% to estimate the [illegible]pected percentage of correct predictions [illegible] Group HR, the 68% confidence interval [illegible] between 83.9% and 99.5%.

TABLE 2

SUMMARY OF PROBLEM-SOLVING RESULTS

Task	Experimental (HR) *Ss*												Control *Ss*												Disqualified *Ss*						
	8 G	8 G	9 G	7 B	8 B	8 B	8 B	8 G	8 G	7 G	8 B	9 B	8 B	7 G	8 B	7 G	G	7 B	8 G	9 B	9 B	8 B	8 G	7 B	8 G	9 B	8 B	8 B	8 G	9 B	5 G
Interpreting single rule cards	+	+	+	−	+	+	+	+	−	−	−	+	−	−	+	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+
Interpreting composite rule cards	+	+	+	−	+	−	+	−	−	+	−	+	−	+	−	−	+	−	+	−	+	+	+	+	+	+	+	+	+	+	−
Pretest		−	−	−			−		−	−	−	−			−				−	−		−	−	−	+	+	+	+	+	+	−
Minutes on experimental HR or irrelevant instruction	6	5	4	5	3	4	7	3	4	4	5	4	[illegible]	5	4	5	3	4	7	4	4	4	5	3	5	5	5	5	5	5	
Transfer (Re) test	+	−	−	[illegible]	−	+	+	+	−	+	−	−	−	−					−	−		−		−	+	+	+	+	+	+	−

Note. The + indicates that *S* reached criterion [illegible] that he did not. The *Ss* are identified according to age (5, 7, 8, and 9) and sex (B, G).
a Upon readministration of the HR treatment, this *S* required 2 min of instruction [illegible] reached criterion.

EXPERIMENT II

The purpose of this experiment was to test the mechanism proposed above with two different and more complex HO rules of varying generality. Both involved generalization from a restricted rule to one that was more general. To help make it easier to identify common and disparate features, the method description parallels that of Experiment I.

Method

Tasks and materials. The experimental tasks involved responding with appropriate numerals to given stimulus numerals. Each task can be charac-

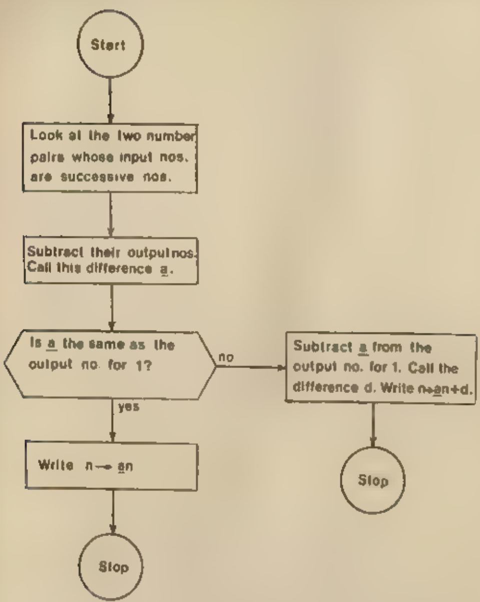


FIGURE 1. Flow diagram for the higher order finite differences rule which acts on restricted rules and generates rules of the form $n \rightarrow an + d$.

terized as a set of stimulus–response pairs in which each number n is mapped into a number of the form $an + d$, where a and d are whole number constants (e.g., n into $4n + 13$). In addition to the above, the domains of the *restricted* tasks contained exactly three numbers, 1 together with (any) two consecutive numbers (e.g., 8 and 9). Restricted tasks were represented as triples which also designated the underlying rules. Triples were always presented to *S*s together with the three input numbers; the pairs were printed on the left of a sheet of paper and the inputs on the right. The rules underlying the *unrestricted* tasks were of two types. The an rules simply involved multiplying the input numbers by a (i.e., $d = 0$). The $an + d$ rules involved both multiplication and addition (of d).

Two HO generalization rules were identified. Both HO rules act on restricted rules (triples) and generate output rules of types an and $an + d$, respectively. A flow diagram for the more general finite differences (F) rule is given in Figure 1. The division (D) rule is similar. Both HO rules partition the class of restricted rules into two equivalence classes consisting of an rules and $an + d$ rules, respectively. The D rule only generates an rules, whereas the F rule generates $an + d$ rules. (The D rule may also be viewed as having a limited domain consisting of only those restricted rules in which the output of 1 equals the quotient of the output in one of the other pairs divided by its input.)

Subjects, design, and procedure. The 80 *S*s were students in Grades 5–12 of Catholic and public schools of Philadelphia. They were run individually. Eight *S*s were eliminated because they were unable to do simple arithmetic computations. The experiment was also terminated early (to ensure equal n in experimental groups and save *E* time) with 12 more *S*s who were able to solve an problems but not $an + d$ problems on the pretest (Phase 3 below).

Experiment II was also run in the following five phases:

1. *S* was presented with triples of number pairs and shown how to interpret them. He was instructed "to write the output that goes with each input according to the triple." The criterion was two correct solutions in succession.

2. The *E* introduced rules of the form $n \rightarrow an + d$, gave one example of a general rule where $d = 0$ and one where $d > 0$, and stated that for every triple there is a rule of the form $n \rightarrow an + d$, which "fits" the triple. The *S* was told: "A general rule fits a triple if, when you use the rule on the input numbers...you get the same outputs as in the triple." The *S* was taught how to check whether a rule fits a triple and practiced checking with four different rules, two $an + d$ and two an. Then *S* was presented in turn with 6 pages, each containing a rule, a triple, and two–four inputs not in the triple. He was told to "find what output goes with each new input number according to the rule." This involved first checking to determine whether the given rule fits the triple and, if so, to use it to compute the output for each new input. The criterion was four consecutive correct problems.

Just prior to the pretest *E* summarized for *S* what he had learned. That is, there are two ways to find the output for a given input of a triple: (*a*) look at the triple to see what the output for that input is, and (*b*) use a rule, which fits the triple, to compute the output for the input. The *S* also was reminded how to check whether a rule fit a triple; and, that, given a triple and a rule that fits the triple, the outputs for new inputs can be computed via the rule.

3. The *S* was given a pretest consisting of four randomly sequenced transfer problems, two an and two $an + d$. Each problem consisted of a triple and three new inputs (but no rule). The *S* was instructed to "find out what output goes with each new input according to a rule...which fits the triple." If *S* solved one but not both problems of Type an ($an + d$), he was given another problem of that type. If he solved it, he was given the unsolved problem again. Under these conditions, if *S* solved the original two problems, he was considered able to solve that type in general, otherwise not. When *S* encountered difficulty on the pretest, *E* explained that the task was difficult and provided encouragement. Based on pretest results, *S*s were categorized into three classes: none—unable to solve either type of problem; AN—able to solve an but not $an + d$ problems; and AN + D—able to solve both an and $an + d$ problems. The none class was randomly divided into three groups: C—*S*s receiving no HO rule training; D—*S*s taught the D rule; and F—*S*s taught the F rule. Class AN was

rand[illegible]ided into Treatment Groups D and F. The AN + D Ss effectively knew the identified HO rules and were unassigned.

4. Each S assigned to Groups D or F was told, "I will teach you a procedure which can be used for some kinds of triples to find a general rule... which fits the triple." On each HO task S was presented with a triple and required to construct a rule that fit the triple. A description of the appropriate HO rule was placed on a stand in front of S, and E worked through two HO practice tasks with S. Then S was presented in turn with six more HO tasks and required to reach a criterion of four correct solutions in a row. The Classes C and AN + D Ss (the latter solved all pretest problems) received no treatment. Just prior to the posttest E again summarized what S had learned, adding for Ss in the HO treatment groups: "You know a procedure which can be used for some kinds of triples to find a general rule... which fits the triple... refer to this procedure whenever you wish." The HO rule remained in front of S during the posttest.

5. The posttest involved new transfer problems but paralleled the pretest in every respect.

Results

Of 60 Ss admitted, 50 were placed in the five experimental groups; 10 passed both the *an* and *an* + *d* problems on the pretest (and the posttest). The experimental results for the 50 Ss were exactly as predicted. No S in Group C of the none class was successful on either the *an* or *an* + *d* posttest problems. All 10 Ss in Groups D and F, of Classes AN and AN + D, were successful on the AN problems of the posttest and all Ss in Group F of both classes were successful on the *an* + *d* problems.

In short, Ss were successful on posttest problems if and only if they knew how to solve the problems before entering the experiment (given only the restricted rules) and/or they were taught an appropriate HO rule.

Experiments III and IV

Experiments III and IV were conducted to determine the feasibility of predicting the performance of individual Ss on specific problems by assessing Ss' knowledge relative to the HO generalization rules identified. Specifically, could the methods developed in Scandura (1973) be used to determine which parts of the HO rules used in Experiment II were available to individual Ss? And, could this information be used to predict the performance of individual Ss on new problems?

Method

The tasks and rules were as in Experiment II. The materials were six booklets. Booklets 1–3 covered the same material as Phases 1 and 2 of Experiment II. Material for Phase 3, pretest on criterion problems, and Phase 4, training on HO rules, were eliminated. Booklets 4 and 5 were substituted for this material. These booklets made it possible to assess knowledge relative to the HO rules. Because outputs of HO rules are themselves rules, Booklet 4 was used to provide instruction in how to write rules of the form $n \to an + d$. Page 1 read, "to write a rule... in which we multiply the input number by 3 and add 4 to the product, let n be the input number. To multiply n by 3 write $3 \times n$. To add 4 to the product write $3 \times n + 4$." Eight practice problems were given. Booklet 5 tested for the HO rules, "On each page . . . is a triple. Find and write . . . a rule of the form $n \to an + d$ which fits the triple." Four problems followed. Two problems required use of an HO rule for deriving $n \to an$ rules and two problems for deriving $n \to an + d$ rules. Booklet 6 corresponded to Phase 6, the posttest.

Experiment III was run with 17-first year algebra students and Experiment IV, with 9 general math and 11 algebra students, all from West Philadelphia. In Experiment III, Booklets 1–6 were administered during a 1-hr. class period. In Experiment IV, Booklets 1–4 were administered on one day, and a short review booklet and Test Booklets 5 and 6 on the next. Both class periods were 40 min. long. In contrast to Experiment II, instructions were read to classes instead of individuals. Each S had to work at least four problems correctly in each booklet. From 3–5 proctors were available in each classroom. In scoring Test Booklets 5 and 6, problems were divided into two categories, according to whether an *an* + *d* or an *an* rule was involved. If S got both problems of one type in a booklet correct, he was considered successful; otherwise not. Although Tests 5 and 6 both involved HO rules, the difference between them is critical. Predictions were based on the prior availability of appropriate HO rules. An S who could derive *an*, but not *an* + *d*, rules in Booklet 5, for example, was assumed able to solve only transfer problems involving *an* rules.

Results

The results of Experiments III and IV are summarized in Table 3. In all cases, the number of correct, as opposed to incorrect, predictions differed significantly from chance ($p < .05$ for both *an* and *an* + *d* problems in Experiment III; $p <$

TABLE 3

SUMMARY OF GENERALIZATION ASSESSMENT RESULTS

Group	No. successful on assessment test (5)	Proportion correct predictions on transfer test (6)	No. of *S*s *not* successful on assessment test (5)	Proportion correct predictions on transfer test (6)	Total no. of correct predictions	Overall percent correct predictions
Experiment III ($Ns = 17$)						
an problems	7	4/7	10	10/10	14	82
an + *d* problems	2	2/2	15	13/15	15	88
Experiment IV ($Ns = 20$)						
an problems	13	11/13	7	7/7	18	90
an + *d* problems	3	2/3	17	17/17	19	95
Combined ($Ns = 37$)						
an problems	20	15/20	17	17/17	32	86
an + *d* problems	5	4/5	32	30/32	34	92

Note. The numbers in parentheses refer to Booklets 5 and 6.

.001 and $p < .05$, respectively, for *an* and *an* + *d* problems in Experiment IV; and $p < .001$ and $p < .005$, respectively, for *an* and *an* + *d* problems in the combined study; exact probability, Finney, 1948). Although the difference was not reliable, the additional 20 min. provided in Experiment IV appeared to increase precision of prediction on both *an* and *an* + *d* problems (8% and 7%, respectively). Whether or not level of prediction could be further increased is not clear, but this observation tends to support the notion that results of such experiments may be expected to conform to prediction just to the extent that memory-free conditions are realized.

DISCUSSION

Overall, the results of Experiments I and II provide strong support for the postulated mechanism. When the effects of memory are minimized, and *S*'s goals are known to *E*, availability of appropriate higher and lower order rules appears to be both a necessary and sufficient condition for solving transfer problems. Although its importance was not emphasized, ability to determine whether or not a given or derived rule satisfies a higher level goal is also crucial (Scandura, 1973). The results of Experiments III and IV show further how availability of HO rules may be determined through testing and used as a basis for explanation and prediction in problem solving. These results, however, apply only in situations where memory and problem definition are not likely to be involved. Extension of the proposed mechanism to include memory and to some extent problem [illegible]inition is given in Scandura (1973). Empiri[illegible] testing should be a first order of business.

Another feature of this research [illegible]h deserves mention is the possibility of [illegible]matically identifying the basic rules un[illegible]ing a set of problems (cf. Scandura, [illegible]n, & Wulfeck, 1972). In general, the [illegible] (including HO rules) underlying a set o[illegible]blems may be determined as follows: (*a*) [illegible]ect a broad (finite) sample of problems in [illegible]he set and identify solution rules for each problem. These rules provide a sufficient basis for solving not only the sampled problems but all other problems "like" them. The use of such rules corresponds to *reproductive* problem solving (Wertheimer, 1945) and, aside from relative complexity, to performance theories in computer simulation (e.g., Newell & Simon, 1972). (*b*) Identify parallels among the identified rules (e.g., they may involve successive application of simple trading rules, or they may be of the form (*an* + *d*). These "parallels" indicate the presence of higher (and lower) order rules from which the solution rules may be constructed. The composition and simple trading rules of Experiment I (from which the composite rules may be derived) and the generalization and restricted rules of Experiment II (from which the *an* and *an* + *d* rules may be derived) provide examples. This type of analysis (of rule sets) may be repeated as many times as desired. In general, the more basic rule sets obtained in this way make it possible, according to the postulated mechanism (Scandura, 1973), to solve a set of problems larger than originally envisioned and correspond to *productive* problem solving (Wertheimer, 1945).

REFERENCES

BOURN , JR., EKSTRAND, B., & DOMINOWSKI, R. *logy of thinking.* Englewood Cliffs, N.J. tice-Hall, 1971.

DUNC . On problem-solving. (Trans. by L. es) *Psychological Monographs,* 1945, 58 No. 270).

DUR , & SCANDURA, J. M. An algorithmic appr h to assessing behavior potential: Comparis with item forms and hierarchical technol s. *Journal of Educational Psychology,* 197 65, 262–272.

FINNE , D. J. The Fisher-Yates test of significance in × 2 contingency tables. *Biometrika,* 1948, 35, .5–156.

GA , R. M. *The conditions of learning.* (2nd ed.) N York: Holt, Rinehart & Winston, 1970.

GRE NO, J. G. The structure of memory and the ress of solving problems. In R. Solso (Ed.), *ntemporary issues in cognitive psychology: Loyola Symposium on Cognitive Psychology.* New York: Wiley, 1973.

NEWELL, A., & SIMON, H. A. *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall, 1972.

SAUGSTAD, P. Problem solving as dependent on availability of functions. *British Journal of Psychology,* 1955, 46, 191–198.

SCANDURA, J. M. Role of rules in behavior: Toward an operational definition of what (rule) is learned. *Psychological Review,* 1970, 77, 516–533.

SCANDURA, J. M. *Structural learning: I. Theory and research.* New York: Gordon & Breach, 1973.

SCANDURA, J. M., DURNIN, J. H., & WULFECK, W. H., II. Higher order rule characterization of heuristics for compass and straightedge constructions in geometry. *Artificial Intelligence,* in press.

WERTHEIMER, M. *Productive thinking.* New York: Harper, 1945.

(Received September 1, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 992–1000

MEMORY SEARCH OF CATEGORIZED LISTS: A CONSIDERATION OF ALTERNATIVE SELF-TERMINATING SEARCH STRATEGIES [1]

MARY J. NAUS [2]

Wellesley College

The two three-stage memory search models for categorized lists described by Naus, Glucksberg, and Ornstein were investigated in three experiments. Both models assume that memory search is serial and terminates after the probed category has been exhaustively searched. The models differ only in whether memory entrance is assumed to be directed to the probed category or to occur randomly. The latency data in Experiment I support random entrance for two-, three-, and four-category lists. When trained in directed entrance, *S*s in Experiment II showed that along with the greater reduction in search time they obtained with random entrance, additional intercept time is required to locate the probed category in the list. Results from Experiment III indicated that *S*s only use random entrance when the categories in memory are blocked. Sternberg's exhaustive model for single-category lists is proposed to be a special case of this three-stage model.

Sternberg (1966) proposed that the retrieval of information from short, single-category lists in recognition memory is serial and exhaustive. Since memory search is obviously not always exhaustive, what are the characteristics of this exhaustive search?

The task in which Sternberg provides evidence for an exhaustive memory search requires *S* to memorize a short set of items (called the *positive set*). A test item (*probe*) is then presented and *S* is asked to decide if the probe was a member of the positive set. If the probe is a member of the set, a positive or *yes* response is correct; if not, a negative or *no* response is required. Reaction time (RT) is measured for these *yes* and *no* responses. Memory search was considered exhaustive because the slopes for the positive and negative responses were equal. Recently, however, there is some evidence that an exhaustive model of memory search may not adequately describe information retrieval in this task. When the items in the positive set are not members of the same category but can be grouped perceptually (Williams, 1971), syntactically (Clifton & Gutschera, 1971), semantically (Naus, Glucksberg, & Ornstein, 1972), or symbolically (Milles, 1969), *S*s are able to use the categorization to modify the exhaustive nature of their memory search.

In an attempt to precisely describe the search of multicategory sets, Naus et al. (1972) proposed two nonexhaustive search models. Both of these models consider memory search to be a three-stage process rather than the single-stage one Sternberg (1966) implied. Both models assume that (*a*) memory entrance includes both the selection of the to-be-searched categories from the positive set and the transfer of these categories into short-term memory for search (*b*) search is serial, and (*c*) search terminates as soon as the probed category has been exhaustively searched. The two models differ only in their rule for the selection of the to-be-searched categories from the set. In the Directed Entry Model, the category of the probe is used both

[1] This research was submitted in partial fulfillment of the requirements for the PhD degree at Princeton University. Appreciation is expressed to the members of the dissertation committee: Sam Glucksberg, Chairman, Peter Ornstein, Carl Sherrick, Tom Trabasso, and Walter vom Saal for their kind encouragement and many helpful suggestions. This research was supported by U. S. Public Health Service Predoctoral Fellowship 5-F01-MH-44648-02 and by Grant MH-10742 administered through Princeton University with Sam Glucksberg as principal investigator.

[2] Requests for reprints should be sent to Mary J. Naus, Department of Psychology, Pendleton Hall, Wellesley College, Wellesley, Massachusetts 02181.

to *begin* and to terminate search, such that the number of items searched equals the number of items in the probed category. The Random Entry Model proposes that the categories entered for search are independent of the category membership of the probe. If the items searched first are from the probed category, search terminates after this category has been exhaustively searched. If the items searched first are from a nonprobed category, search continues until the probed category has been completely searched. When Naus et al. (1972) presented *S*s with sets of both one and two semantic categories, the two-category sets showed a slope reduction of 24.8%. This is consistent with the Random Entry Model's predicted 25.0% slope reduction for two-category sets.

The experiments presented here further investigate memory search of multicategory sets. Experiment I extends the number of semantic categories in the sets to three and four to determine whether *S*s continue to use a random entrance process. Experiment II is designed to teach *S*s to use the Directed Entry Model in order to specifically compare the results of a directed vs. random search. Finally, whereas in the first two experiments the categories in the sets were presented in blocked order, Experiment III presents them in random order to determine the generality of random entrance.

Experiment I

The Random and the Directed Entry Models make differential predictions for the number of items searched as a function of the number of categories in the set (K). For the Random Entry Model the predicted percent reduction in multicategory slope equals $[(1.00 - (K+1)/2K)\cdot(100)]$, while for the Directed Entry Model the predicted percent reduction in slope for multicategory sets is $[(1.00 - 1/K)\cdot(100)]$ (Naus et al., 1972). This experiment was designed to determine whether *S*s continue to use random entrance for three- and four-category sets.

Method

Experimental design. A mixed experimental design with three groups of *S*s was used. Group 1C/2C was presented sets of 2, 4, 6, and 8 words, with half of the sets containing words from one category and half containing words from two. Group 1C/3C received sets of 3, 6, and 9 words, with half of the sets containing words from one category and half containing words from three. Finally, Group 1C/4C saw sets of 4, 8, and 12 words, with half of the sets containing words from one category and half containing words from four. Each *S* received all combinations of positive set sizes, response types, and number of categories for his group.

In each multicategory set there was an equal number of words from each category. This restriction, together with the different number of categories in each group, yielded the different set sizes across the three groups. The categories in the multicategory sets were presented in blocked order. Half of the tests were negative and half were positive. Negative probes were only selected from categories represented in each set. The sequence of set sizes, the number and order of categories, the positive or negative test trials, and the serial position of the probes were randomized and counterbalanced over trials.

Stimulus materials. The total ensemble of stimulus words varied across the three groups: Group 1C/2C had a total ensemble of 32 words, 16 animal and 16 girls' names; Group 1C/3C had a total of 48 words, adding 16 musical instrument names to the ensemble for Group 1C/2C; and Group 1C/4C had a total of 64 words, adding 16 vegetable names to the ensemble for Group 1C/3C. The instances of these four taxonomic categories were obtained from the Battig and Montague (1969) category norms, equating the categories for "cohesiveness" based on the mean rank of the words eliciting the categories. The instances of these categories were also equated for mean frequency, number of syllables, and length on the basis of the Thorndike and Lorge (1944) word frequency count. The stimulus words were made into slides and were also printed on index cards for presentation in the free recall learning.

Apparatus. A Kodak Carousel projector, equipped with a Lafayette tachistoscopic shutter, displayed the slides on a translucent screen between the control room and *S*'s Industrial Acoustic Company chamber. Beneath the screen in *S*'s room were a "Y" indicating the position for the *yes* response and an "N," the position for the *no* response. The *S* sat at a small table facing the screen. On the table was an intercom connecting *S*'s room with the control room and a response box containing a small ready light, two microswitch response keys, and two feedback lights. The feedback lights were located 5 in. behind each key. The ready light was mounted between the two response keys. The response keys were connected to a digital klock-

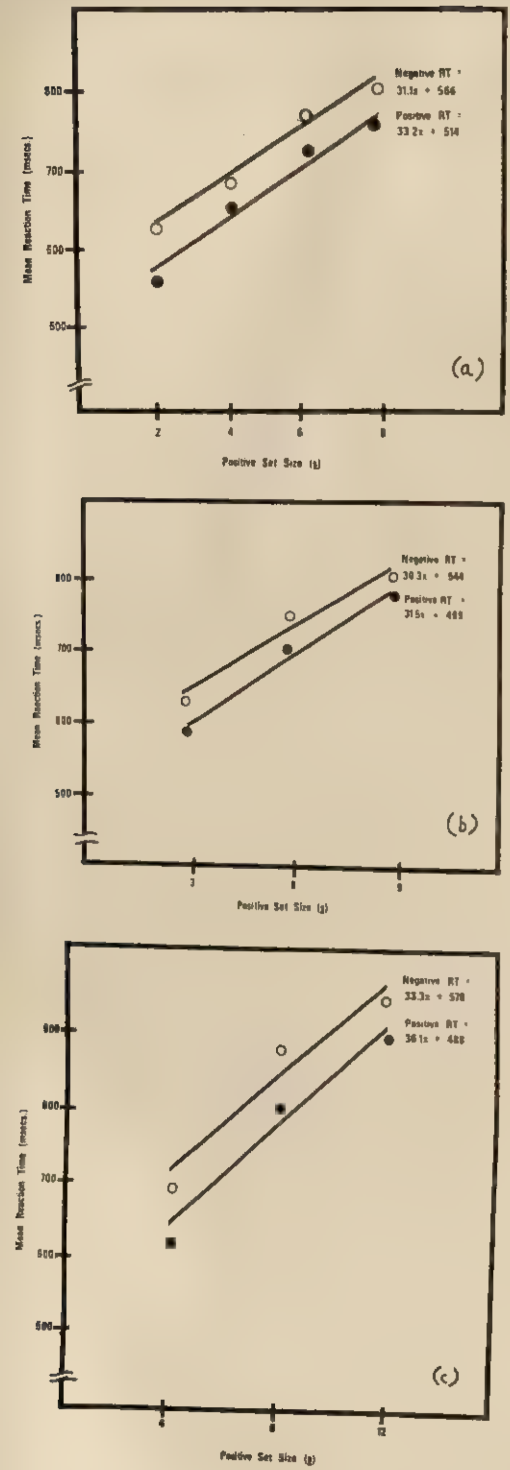


FIGURE 1. Mean reaction time (RT) as a function of positive set size for one-category sets for (a) Group 1C/2C: Experiment I, (b) Group 1C/3C: Experiment I, and (c) Group 1C/4C: Experiment I.

kounter and two signal lights to i e to *E* which key was pressed. The projector, r, and ready and feedback lights were controll Tally Tape Reader. All presentation interva l light durations were controlled by electro ical timers.

Procedure. The procedure wa same for all three experimental groups. In ea ur sessions, *Ss* were initially shown the stimul rds on cards and were instructed to memori recall the words in any order. At least 90 call was required. The *S* then received t recognition–memory instructions used by Stern (1966), and was also told that the categorized rds he had just memorized would be used in t e sets. Each trial of the RT task followed a fication of Sternberg's fixed-set procedure, wit set changing after every eight tests.

The number of tests per 60-min ion varied across the groups: Group 1C/2C re l 280 tests per session, Group 1C/3C received nd Group 1C/4C, 224.

Subjects. Eighteen right-handed me om Princeton University were paid $10 each t as *Ss*.

Results and Discussion

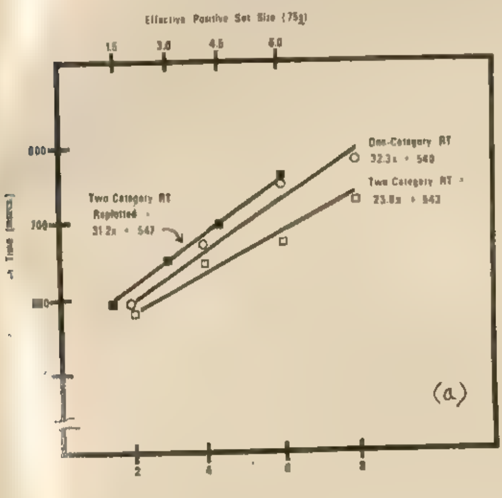
Since primary interest was in mptotic, error-free performance, only sion IV data for correct responses are eported. The error rates were between .9% and 2.2% for both the one- and mul ategory sets for all three groups. These e ror rates did not significantly differ from on another. There were also no differences in error rates for positive and negative respons

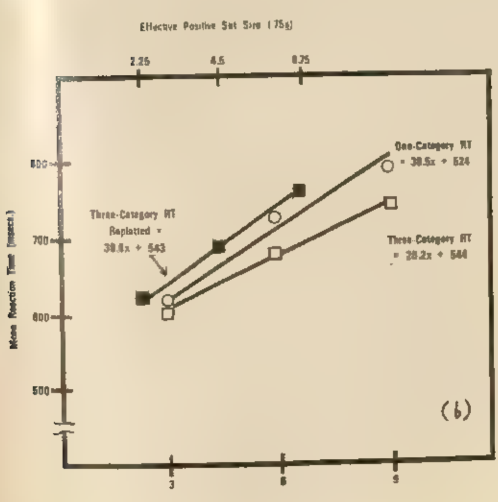
One-category results. The one-category results were consistent with those reported by Sternberg (1966) for all three groups. These data are summarized for Groups 1C/2C, 1C/3C, and 1C/4C in Figure 1a, 1b, 1c, respectively. Mean RT increased significantly with positive set size, the positive and negative slopes were equal, and the positive RTs were reliably faster than the negative RTs.

Category effect: Means and variances. To assess the effect of taxonomic categories on RT, the one- and multicategory sets in each group were submitted to separate analyses of variance. For all groups, the mean RTs for multicategory sets were reliably faster than the mean one-category RTs, indicating that *Ss* used the categories to modify their search strategies. Specifically, these one- and multicategory differences in mean RT were a decrease in the multicategory slopes, as reflected by significant Number

of Categories × Positive Set Size interactions for each group.

To determine if *S*s were specifically using the Random Entry Strategy, the slopes and intercepts for the one- and multicategory sets were directly compared. These comparisons were applied to data pooled over positive and negative trials because there were no reliable Positive–Negative × Categories interactions. These data are summarized in Figure 2a, 2b, and 2c. When the one- and multicategory data are based upon the size of the set actually presented (*s*) the one- and multicategory slopes differed reliably from one another for all of the groups: for Group 1C/2C, $t\ (5) = 8.86$, $p < .01$; for Group 1C/3C, $t\ (5) = 12.80$, $p < .01$; and for Group 1C/4C, $t\ (4) =$





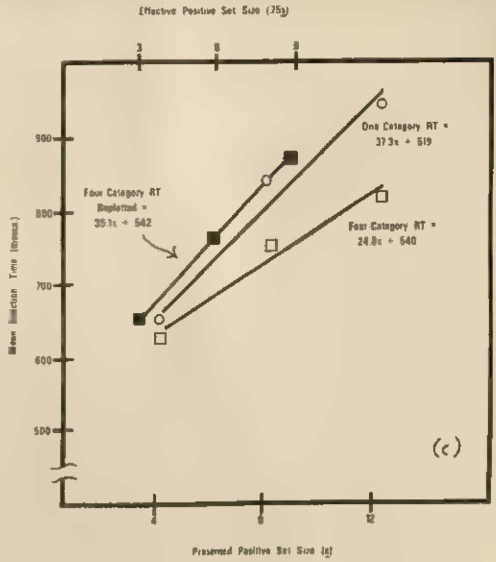


FIGURE 2. Mean reaction time (RT) as a function of presented positive set size for (a) one- and two-category sets and mean RT as a function of effective positive set size for two-category sets: Experiment I, (b) one- and three-category sets and mean RT as a function of effective positive set size for three-category sets: Experiment I, and (c) one- and four-category sets and mean RT as a function of effective positive set size for four-category sets: Experiment I.

15.83, $p < .01$. The obtained percent slope reduction was 26.3 for two-category sets, 33.8 for three-category sets, and 35.7 for four-category sets. These obtained percent slope reductions are consistent with the predictions of the Random Entry Model.

While the average data are consistent with the Random Entry Model, it is important to determine if this model was used by the majority of *S*s. In Group 1C/2C, four of the six *S*s had one-category–two-category slope differences which were within the 95% confidence limits of a 25.0% slope difference. In Groups 1C/3C and 1C/4C only S_6 in Group 1C/4C did not show results consistent with the Random Entry Model.

Further evidence in support of the Random Entry Model may be obtained from the distribution of RTs. Since the Random Entry Model asserts that the RTs for multicategory sets are an additive combination of the time required to search proportions of the set ranging from one to

K categories, the multicategory variances can be predicted based upon the obtained one-category variances for sets of different sizes. The equation for predicting two-category variances presented by Naus et al. (1972) can be extended to three- and four-category sets. When an F test was used to compare the obtained and predicted multicategory variances, no differences were obtained for any of the groups.

Category effect: Intercepts. Since the number of items actually presented in the sets does not represent the number of items actually searched given random entrance (Naus et al., 1972), the multicategory sets have been replotted as a function of *effective* set size (see Figure 2a, 2b, and 2c). Given that these replotted slopes and the one-category slopes do not differ, the intercepts can be directly compared. Although the multi- and one-category intercepts did not reliably differ for any of the groups, the magnitude of the difference seemed to be a direct function of the number of categories in the set: on the average 7 msec. for two-category sets, 22 msec. for three-category sets, and 34 msec. for four-category sets. A linear regression analysis on the intercept differences as a function of the number of categories in the set yielded a slope of 13.5 which significantly differs from zero, $t\ (30) = 2.14$, $p < .05$. This increase in multicategory intercept as a function of the number of categories in the sets implies that there is an additional constant operation involved in a Random Entry Search which is directly related to the number of categories in the set. This operation could be the en bloc transfer of each of the to-be-searched categories to short-term memory for search. This proposed transfer process suggests a possible mechanism for the random selection of at least the first category searched: the category in short-term memory when the probe is presented is the category that is searched first. Since the category in short-term memory when the probe is presented has the probability of $1/K$ of being the probe category this hypothesis predicts that the Random Entry Model requires $\sum_{i=0}^{K-1} i/K$ transfers, each costing 13.5 msec. in intercept time. The multicateg[illegible] intercepts predicted by this hypothesis[illegible]wever, are smaller than those actuall[illegible]tained for three- and four-category s[illegible]suggesting that another "constant" op[illegible]ion is also involved.

Given that the Directed [illegible] Model is more efficient in reducing t[illegible]umber of memory comparisons and th[illegible]s seem to use a directed entry search f[illegible] large categorized sets (Okada & Burrow[illegible]1973), why do Ss use the Random Ent[illegible] Strategy? The data from S_6 in Group 1[illegible]C suggest that a thorough comparison of the two models may provide an ans[illegible] to this question. This S's mean da[illegible]upport a directed search. In addition, [illegible]ifference between the one- and four-ca[illegible] intercepts, where the four-category[illegible]ercept is determined as a function of [illegible] Directed Entry Model, is 130 msec[illegible]hus, the Directed Entry Model seems [illegible]chieve a larger reduction in the numb[illegible] items searched only at the cost o[illegible]dditional intercept time, such as the ti[illegible]equired to select the probe category f[illegible] the set prior to search.

Experiment II

Method

Experimental design. The Ss were either in Grou[illegible] 1C/3C or 1C/4C. These groups were divided into the regular and the directed-entry instruction conditions. The regular instruction Ss were given the same instructions as in Experiment I. The directed entry Ss were instructed and taught to use the Directed Entry Model. The instructions stressed directed selection of the probed category prior to search onset. To aid Ss in using the Directed Entry Model, they were also provided with an auditory presentation of the probe category prior to the presentation of the probe. This auditory category cue occurred simultaneously with the onset of the warning signal, about .8 sec. prior to the presentation of the probe.

Procedure. Initially both instruction conditions received the free recall task and the recognition memory instructions described in Experiment I. In addition, for the first three sessions, Ss in the directed-entry instruction condition received instructions that stressed the direct selection of the probed category prior to search onset and that informed them that the probe category would be auditorily presented prior to the presentation of the probe to aid in their memory search of only the probed category. Session IV did not differ for directed-entry and regular instruction Ss.

Subjects. [T]welve right-handed undergraduate men from [Prin]ceton University were paid $10 each for servi[ng as] [S]s.

Results [an]d Discussion

Bot[h] [S]essions III and IV data are considere[d] [b]ecause of the differential treatment in th[e] [d]irected-entry instruction condition. The [e]r[r]or rates for these sessions for one- and [mu]lticategory sets ranged 1.9%–2.7% for [bot]h positive and negative trials. These err[or] [r]ates did not reliably differ for any of th[e] [ce]lls of the design.

[One-]category results and category effect for re[gu]lar instructions. The one-category re[s]u[lts] for both groups and instruction c[on]ditions supported those reported by [Ste]rnberg (1966). The regular instruction [con]dition was a control condition and repli[ca]ted Experiment I, showing a slope reduction of 33.9% for Group 1C/3C and 38.0% for Group 1C/4C.

Category effect: Directed-entry instructions. In order to determine if *S*s in the directed-entry instruction condition used the Directed Entry Model to search the multicategory sets, separate linear regression analyses were conducted for Groups 1C/3C [a]nd 1C/4C for Sessions III and IV. Session [I]II data summarize the effect of directed-[en]try training while it occurred. The ob[tai]ned one- and multicategory slopes dif[fe]red reliably for each group: for Group 1[C/]3C, t (2) = 36.6, $p < .01$, and for Group 1C/4C, t (2) = 39.6, $p < .01$, with the obta[i]ned slope reductions of 70.8% for Group 1C/3C and 76.6% for Group 1C/4C consistent with a Directed Entry Search. Individual *S*'s slope reductions were also consistent with the Directed Entry Model. The variance data, when the predicted multicategory variances are based on the obtained one-category variances for Set Size s/2, were also consistent with directed entrance.

During Session IV, *S*s no longer received instruction or training in the use of the Directed Entry Model. Although the obtained one- and multicategory slopes differed reliably for both groups—for Group 1C/3C, t (2) = 9.4, $p < .01$, and for Group 1C/4C, t (2) = 39.11, $p < .01$—there was a difference in obtained slope reductions between the groups. The obtained slope reduction of 34.0% for Group 1C/3C was consistent with the Random Entry Model, whereas the obtained slope reduction of 76.4% for Group 1C/4C supported a Directed Entry Search. For both groups, individual *S*'s slope reductions were consistent with these averaged results and the variance analyses supported the findings obtained with mean data. Thus, in contrast to Group 1C/3C, where *S*s returned to the use of the Random Entry Model when training and instruction were discontinued, Group 1C/4C *S*s continued to use the Directed Entry Model during Session IV.

Direct comparison of the one- and multicategory intercepts again requires that multicategory intercepts be reestimated as a function of effective set size based upon either directed or random entrance. The one-category and reestimated multicategory data are summarized in Table 1. Since there were no reliable differences between one-category and reestimated multicategory slopes, the intercepts can be directly compared. When *S*s used the Directed Entry Model, there was a large increase in multicategory intercept: on the average 140 msec. for three- and 185 msec. for four-category sets.

Comparison of the models. An in-depth comparison of these models may answer two important questions. Why do *S*s not always use the Directed Entry Model to search the multicategory sets since it leads to a greater reduction in memory comparisons, and why is there a differential effect for Groups 1C/3C and 1C/4C during Session IV of this experiment?

The Directed Entry Model leads to a large increase in multicategory intercepts as well as to a large reduction in the number of memory comparisons. Two processes seem to be included in these multicategory intercepts. First, the Directed Entry Model requires time to transfer the probed category to short-term memory. Assuming that this transfer only occurs if the probed category is not already in short-term memory when the probe is presented, on the average $1 - 1/K$ transfers are required. Second, the additional intercept time could

TABLE 1

AVERAGED DATA FOR THE DIRECTED-ENTRY INSTRUCTION CONDITION WITH THE MULTICATEGORY PARAMETERS ESTIMATED AS A FUNCTION OF EFFECTIVE SET SIZE AS PREDICTED BY EITHER THE RANDOM OR DIRECTED ENTRY MODELS

Group and no. of categories	Session III				Session IV			
	Intercept (msec.)	Slope	Intercept difference (msec.)	Absolute slope difference	Intercept (msec.)	Slope	Intercept difference (msec.)	Absolute slope difference
1C/3C								
1C	409	41.5	141*	.1	494	24.7	19	.7
3C	550	41.4			513	25.4		
1C/4C								
1C	401	47.0	183**	1.4	407	45.4	187**	.3
4C	548	45.6			594	45.1		

* $p < .01$.
** $p < .001$.

be the time required to select the probed category from the set prior to search. If this category selection is considered to be a serial matching operation between the category of the probe and the categories in the set, this operation would be represented by an increase in intercept which is a direct function of the number of categories in the set, approximately 45 msec. per category. Although Williams (1971) has found evidence for a serial and exhaustive category comparison process for perceptual categories, the present experiment does not indicate whether the category comparison process for semantic categories is exhaustive or self-terminating. Together, these two operations add $(1 - 1/K) \cdot 13.5 + (K) \cdot 45$ msec. to the one-category intercept resulting in a predicted three-category intercept increase of about 145 msec. and a four-category intercept increase of about 190 msec. These results support a net-savings hypothesis (Naus et al., 1972) which suggests that *S*s spontaneously use the Random Entry Model because in this experiment it is more efficient in *overall* time savings. However, one would suspect that, with additional categories and larger set sizes, the Directed Entry Model would become the more efficient.

In addition to the Random Entry Model's overall time savings in the present experiments, the demands of this experimental procedure, such as temporal sequences and *S*'s instructions, may impose a rapid search onset over selective entry. Together with the net-savings hypothesis this reason might provide an explanation for the differences between Groups 1C/3C and 1C/4C in Session IV. In the regular instruction condition, this task may favor a random search onset for both Groups 1C/3C and 1C/4C, even though the set sizes in the [illegible] condition are large. The training [illegible] in the directed-entry instruction condition modified the task demands sufficiently to favor directed entrance for both groups. However, after training was discontinued, the strategy used to search the multicategory sets depended upon the size and number of categories in the sets.

This experiment raises the question of the generality of the Random Entry Model, suggesting the importance of precise specification of the conditions under which it is used. Experiment III was designed to determine if *S*s will continue to use the Random Entry Model if the category exemplars are presented in random order rather than in the blocked order.

EXPERIMENT III

Method

Experimental design. All *S*s were in Group 1C/2C. The stimulus materials, instructions, and number of trials were the same as in Experiment I with the addition of a postexperimental questionnaire at the end of the last session.

Subjects. Six right-handed undergraduate Princeton males were paid $10 each for their participation in the four sessions of the experiment.

TABLE 2

INDIVIDUAL *Ss*' SLOPE ANALYSES FOR GROUP 1C/2C: EXPERIMENT III

	Obtained one-category slope	Obtained two-category slope	Slope difference		95% confidence interval		Obtained % slope reduction
			Obtained	Predicted	Lower bound	Upper bound	
S_1	37.7	26.7	11.0	9.4[a]	9.2	12.8	29.2
S_2	38.2	27.6	10.6	9.5[a]	8.3	10.7	25.6
S_3	18.9	15.0	3.9	3.7[a]	2.9	4.7	20.6
S_6	31.4	23.8	7.6	7.9[a]	6.9	8.9	24.2
Average	31.3	23.3	8.0	7.8[a]	7.0	8.6	25.6
S_4	32.7	35.5	−2.8	8.2	—	—	—
S_5	33.6	30.8	2.8	8.4	—	—	—
Average	31.2	33.2	2.0	7.8	—	—	—

[a] Consistent with the Random Entry Model's predicted slope difference.

Results and Discussion

One-category results. As in the first two experiments, the Session IV error rates for one- and two-category sets, which ranged 1.0%–.8% did not reliably differ, and the one-category latency data were consistent with those previously reported by Sternberg (1966).

Category effect. When an analysis of variance was completed to assess the effect of taxonomic categories on RT, the mean RTs for two-category sets were reliably faster than the mean one-category RTs, $F(1, 5) = 17.73$, $p < .001$. To determine whether *Ss* were using the Random Entry Model to search the two-category sets, regression analyses were applied to data pooled over positive and negative trials because there was not a reliable Categories × Positive–Negative interaction. The one-category sets had a slope of 31.3 and an intercept of 498 msec. The two-category sets had a slope of 27.8 and an intercept of 505 msec. When the one- and two-category slopes were compared, they did not reliably differ, showing a slope reduction of 11.2%. However, analysis of the postexperimental questionnaire indicated that *Ss* could be divided into two groups. Four of the six *Ss* (*Ss* 1, 2, 3, and 6) reported actively operating upon the random two-category sets to reorganize them into category blocks. Instead of rehearsing the items in the presented order, these *Ss* described reorganizing the positive set items into category groups and rehearsing the items in these blocks. In contrast, the remaining two *Ss* wrote that they did not reorganize the positive set items into groups by category, but memorized the items in their presented order. The individual *S*'s Session IV data with *Ss* grouped as "reorganizers" and "nonreorganizers" are presented in Table 2. The reorganizers one- and two-category slopes differed reliably, $t(3) = 7.23$, $p < .01$, yielding a slope reduction of 25.6%. All individual *S*'s results were also consistent with the Random Entry Model's predictions. In contrast, neither of the nonreorganizers had one- and two-category slopes which reliably differed from one another.

This experiment suggests that memory search strategy selection, at least the use of the Random Entry Model, is influenced by the form of the input. Clifton and Gutschera (1971) showed that presenting two-digit numbers in blocked order, where the category is defined by the tens digit, greatly facilitated the use of a hierarchical search. However, randomizing the digits did not completely eliminate a hierarchical search. The Random Entry Model differs from a digit hierarchical search in that it seems to *require* that the items be stored in memory in blocked categories. Speculatively, this blocked storage might be required for the en bloc transfer of the categories to short-term memory for search. Since all *Ss* did not block the categories in the set in order to use the Random Entry Model, this implies that the selection of a memory search strategy is an option available to *S*, i.e., an *S*-controlled process rather than a structural memory feature (Atkinson

& Shiffrin, 1968). This indicates, as recently demonstrated by Hunt, Frost, and Lunneborg (1973), that individual difference variables may be very important in investigations of memory search.

General Discussion

These experiments suggest that memory search is a three-stage process. When the number of items in memory is small and the items are all members of the same category, results are consistent with Sternberg's (1966) exhaustive search model. This model, however, is proposed to be a special case of the more general three-stage model presented here. The single stage results because with small, one-category sets memory entrance is unnecessary and search always terminates after the complete positive set has been searched. In contrast, when the sets are large or can be organized into categories, memory entrance becomes necessary and search can be nonexhaustive through the use of category boundaries.

The Random Entry Model requires speculation as to how random entrance might occur. It seems reasonable to suppose that the category which is in short-term memory at test is the category which is searched first. It is more difficult to propose a mechanism for the random selection of the remaining categories for search. Random entrance requires that S knows when he has searched a nonprobed category, but does not use this information to select the probed category from the set. It is as if once S is in a "searching mode" it costs more in either time or effort to switch to a "selection mode" to locate the probed category in the set than it does to search additional categories.

In addition to the two nonexhaustive, three-stage search models presented here, the slope reductions for multicategory sets could be interpreted in terms of a faster comparison model. In this class of models, memory search of the multicategory sets is considered to be exhaustive, but some or all of the memory comparisons are proposed to be made more quickly than in the one-category sets. Assume that the members of the probed category are compared to the probe at the one-category rate (x msec. per item), but the items in nonprobed categories can be compared more quickly (y msec. per item). Thus, when the positive set consists of K categories of equal size, the average search time per item is $x + (K-1)y/K$ msec. The decrease in the time required for search of the [illegible]ed category items would be represented [illegible] decrease in multicategory slope. If it is [illegible]her assumed that the items in the nonprob[illegible]ategories are searched at twice the rate as [illegible]tems in the probed category, this model m[illegible]s the same prediction as the Random Ent[illegible]odel. The nonexhaustive models are curr[illegible]ly favored, however, given that the fast[illegible] comparison models cannot adequately ac[illegible]nt for the intercept and variance parame[illegible] in Experiment I and the increased sl[illegible] reductions obtained in Experiment II. E[illegible]imentation is currently under way to direct[illegible]ifferentiate between these two classes of m[illegible]

References

Atkinson, R., & Shiffrin, R. H[illegible] memory: A proposed system and its contro[illegible]esses. In K. Spence & J. Spence (Eds.), *Th[illegible]chology of learning and motivation.* Vol. 2. [illegible]ew York: Academic Press, 1968.

Battig, W. F., & Montague, W. [illegible] Category norms for verbal items in 56 cate[illegible] replication and extension of the Connec[illegible] category norms. *Journal of Experimental P[illegible]ogy Monograph*, 1969, **80** (3, Pt. 2).

Clifton, C., & Gutschera, K. [illegible]ence for hierarchical search processes in [illegible] cognition memory task. *Journal of Verbal [illegible]ning and Verbal Behavior*, 1971, **10**, 528–541.

Hunt, E., Frost, N., & Lunneborg, [illegible]ndividual differences in cognition: A new app[illegible] to intelligence. In G. Bower (Ed.), *The [illegible]chology of learning and motivation.* Vol. 7. New York: Academic Press, 1973.

Milles, K. The effect of a basis for stimulus classification in recognition reaction time. Unpublished master's thesis, Kent State University, 1969.

Naus, M. J., Glucksberg, S., & Ornstein, P. A. Taxonomic word categories and memory search. *Cognitive Psychology*, 1972, **3**, 643–654.

Okada, R., & Burrows, D. Organizational factors in high-speed scanning. *Journal of Experimental Psychology*, 1973, **101**, 77–81.

Sternberg, S. High speed scanning in human memory. *Science*, 1966, **153**, 652–654.

Thorndike, E. L., & Lorge, I. *The teacher's word-book of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

Tulving, E., & Thomson, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–124.

Williams, J. D. Memory ensemble selection in human information processing. *Journal of Experimental Psychology*, 1971, **88**, 231–238.

(Received September 4, 1973)

Journal of Experimental Psychology
1974, Vol. 10 , 1001-1006

DIRECTED FORGETTING AS A FUNCTION OF EXPLICIT WITHIN-LIST CUING AND IMPLICIT POSTLIST CUING[1]

ADDISON E. WOODWARD, Jr.[2]

Governors State University

DENISE CORTIS PARK[3] AND KAREN SEEBOHM[4]

Albion College

Explicit within-list instructions to forget, as well as postlist implicit instructions to forget, were presented within the same list structure to 60 *S*s. The results indicated that explicit forget items were far less likely to be intruded in immediate recall, recalled finally, or recognized on a final recognition test than implicit forget items. These findings support the notion that explicit forget items differ from implicit forget items. Furthermore, while search and retrieval mechanisms may be a viable explanation for phenomena involving implicit forget instructions, differential rehearsal is the most likely explanation for directed forgetting resulting from explicit forget instructions.

The ability of *S*s to differentiate between items that are to be remembered and items that are to be forgotten has merited a significant amount of research (Bruce & Papay, 1970; Elmes, 1969; Epstein, 1969, 1970; Woodward & Bjork, 1971). Various experimental designs have incorporated either implicit (Epstein, 1969, 1970; Shebilske, Wilder, & Epstein, 1971) or explicit (Davis & Okada, 1971; Woodward & Bjork, 1971) instructions to forget in order to investigate directed forgetting in memory. Implicit forget instructions typically ask *S* to exclude some portion of the input from output at the time of recall (Epstein, 1969, 1970). Usually, *S*s are not asked to recall those excluded items, thus the implicit forget label.

Explicit forget instructions are most often incorporated in experiments using item-by-item cuing. The *S*s are presented list items, with each item being followed by either an instruction to remember or an instruction to forget. A final recall test is used to determine the effectiveness of the forget instruction.

Both kinds of paradigms have generated phenomena that have been labeled instances of intentional forgetting or directed forgetting. Presumably, a search for the mechanisms underlying forgetting as a result of implicit forget instructions would also explain forgetting as a result of explicit forget instructions. It is the contention of the present authors, however, that the basic phenomena resulting from implicit or explicit forget instructions are markedly different and, although each paradigm may tell us something about the memory process, neither tells us much about the other.

The present study brings together within a single experiment both implicit and explicit forget instructions. Implicit cuing to forget is defined as any cue that instructs *S*, at the time of output, to disregard part of the input—a situation very similar to the procedures employed by Epstein (1969, 1970). Explicit cuing to forget is defined as any signal or cue, occurring after input but before output, instructing *S* not to remember the item(s) in question. The focus is on the fate of those items with forget instructions at

[1] This research was supported by the Office of Education Grant OEG-5-71-0023(509), and was presented at the meeting of the Midwestern Psychological Association, Chicago, May 1973. The authors extend their gratitude to Robert A. Bjork for his critical comments on an earlier version of this paper.

[2] Requests for reprints should be sent to Addison E. Woodward, Jr., College of Human Learning and Development, Governors State University, Park Forest South, Illinois 60466.

[3] Now at the State University of New York at Albany.

[4] Now at the University of Iowa.

immediate recall, at final recall, and at final recognition.

The explicit cuing is within-list, while the implicit cuing is postlist. Thus, time of cue is being manipulated in a way similar to Bjork's (1970) Experiment III. The implicit forget (F) items in this experiment are similar to items followed by Bjork's remember yellow:forget yellow instruction which urged *S*s first to remember the yellow items and then later to forget those items. Bjork's forget yellow:remember green instruction is equivalent to items followed by an explicit forget instruction. The present experiment differs from Bjork's in that item-by-item cuing was used, the task was free recall, not paired associates, and a final recall test as well as a final recognition test were administered.

Within each of six lists there were three classes of items: items followed directly by a forget instruction, and two remaining sets of items followed by remember instructions but functionally separate. Each list was followed by one of six recall instructions. If one set of remember (R) items is labeled A, the other B, then the recall instructions include recall only A, recall only B, recall A then B, recall B then A, free recall A and B, and finally, after one list, *S*s were asked to recall all the items, even those followed by explicit forget instructions. Every list then contained a set of items followed by explicit forget instructions. Lists that incorporated implicit forget instructions included the recall-only-A instruction (B items were the implicit F items), and the recall-only-B instruction (A items were the implicit F items). At the conclusion of the experiment a final recall test was administered followed by a final recognition test.

The data of interest include the following comparisons: (*a*) the probability of immediate recall, final recall, and final recognition of explicit F items vs. implicit F items (A items from the only-B condition, and B items from the only-A condition); (*b*) the probability of immediate recall, final recall, and final recognition of explicit F items vs. A items and B items in the recall-all condition; (*c*) the probability of immediate and final recall and final recognition of A items in the o[illegible] A condition vs. the probability of recal[illegible] A items in the A then B condition and [illegible] items in the only-B condition vs. B item[illegible] the B then A condition.

METHOD

Subjects. The *S*s were 60 under[illegible]ate students at Albion College. They partic[illegible]d in order to fulfill introductory course requirem[illegible]

Materials and apparatus. All *S*s were presented six 24-item lists. Words were co[illegible]on unrelated four-letter nouns. Eight randomly [illegible]ected words were followed by a red dot, eight b[illegible] a black dot, and eight by a blue dot. In each li[illegible]uarter, two words each were followed by one of th[illegible]olored dots and the same colored dot did not ap[illegible] more than twice consecutively. The distract[illegible]ns used in the final recognition test were selecte[illegible]m the same pool of words as were the target wor[illegible] The words, the cues to remember and forget, [illegible] the recall instructions were presented on a high[illegible]d memory drum (change time less than .05 se[illegible]ated by a prepunched paper tape which determi[illegible] the length of item presentation and recall time[illegible]

Design. All *S*s saw the same lists o[illegible]ms. Each list was followed by a different outpu[illegible]nstruction. The six instruction conditions were r[illegible]l only A, recall only B, recall A and B, recall A t[illegible]n B, recall B then A, and a special list after which *S*s were to recall all items—A, B, and F.

Item sets were designated by th[illegible] different colors. Each word was followed by either a red, black, or blue dot. Hence, there were two sets of remember items and one set of forget items differentiated by color. Each *S* was initially instructed that two colors designated remember items and one color designated forget items.

Counterbalancing techniques were employed so that, across *S*s, each item served in each of the three conditions. This manipulation occurred as a result of changing the instruction associated with a particular color cue from one *S* to the next. Also, across *S*s, each output instruction occurred at each list position with the exception of the recall-all list which was always paired with the third list. Each list began with a 3-sec. ready signal and ended with a 30-sec. recall period. Each item was shown for 3.3 sec. and each instruction for 1.0 sec.

Procedure. The *S*s were run individually, and in the initial set of instructions they were told that they would see lists of words and that each word would be followed by a colored dot which designated whether the item was to be remembered or forgotten. Their task was to try to recall as many of the to-be-remembered items according to the instructions following each list. They were informed that, following each list, there would be one of six possible recall instructions and were told what these instructions would be and that it would be their

best st… to try to remember R items and to forget … and not to try to anticipate when the spec… recall list would occur—a procedure employ… Reitman, Malin, Bjork, and Higman (197… Bjork and Woodward (1973). Such a proced… was deemed successful by Bjork and Wood…ard in that their *S*s did not appear to antic…te the occurrence of the special list. Explicit …rget instructions existed at the onset of the exper…ent when *S*s were told which of the three colo… designated forget items. Implicit forget instru…ns existed for A items in the only-B condition and … B items in the only-A condition. Although the …itial instructions to *S*s indicated what the lik… recall conditions would be, they did not in… which recall instruction was paired with w… list. Since *S*s did not know how many lists w…d be seen, it seemed unlikely that they would b… able to accurately ascertain what a particular …ll instruction would be.

…he *S*s were given three practice lists, each con…ting of 12 two-digit numbers in order to familiarize …hem with the procedure. The recall conditions …aired with the practice lists included recall A then B, recall A and B, and recall only A. After all of the experimental lists had been presented, *S*s were given a final recall test of both R and F items for as much time as they needed. At this point, *S*s were asked to write down each and every word from the experiment that they remembered, regardless of the items' previous instruction. Following final recall, they were then given a final recognition test consisting of the original 144 items and 144 distractors. Their task was to circle any item previously presented to them again regardless of the item's previous instruction. The immediate recall tests, …he final recall test, and the final recognition test …ere all written. The *S*s were instructed in the …dered recall conditions (A then B, B then A) to …rite down as many of the words to be recalled first … they could, to then draw a line and write down … many of the words to be recalled last as they …uld. The *E* attempted to monitor the recall process in those instances providing a partial safeguard against *S*s who wished to circumvent the instruction.

RESULTS

Table 1 presents the probability of immediate recall, final recall, and final recognition of A, B, and F items. Recall of explicit F words is consistently lower than implicit F words. The probability of intruding F items immediately in the only-A condition was .040, while the probability of intruding B items, items that were not to be recalled either, was .119. A correlated t test indicated this difference to be significant, t (59) = 4.12, $p < .001$. In the only-B condition, explicit F items were intruded with a probability of .031, while A items were intruded with a probability of .117. This difference was also significant, t (59) = 4.24, $p < .001$. When *S*s were asked to free recall all items from the list, F items were recalled with a probability of .081, while A and B items were recalled with probabilities of .272 and .254, respectively. A one-way analysis of variance indicated that indeed these differences were real, F (2, 177) = 25.66, $p < .001$.

A 6 × 3 analysis of variance (Recall Conditions × Item Types) on the immediate recall data indicated a significant item effect, F (2, 1062) = 228.36, $p < .001$, a significant effect of recall instruction, F (5, 1062) = 4.98, $p < .001$, and a significant Item Type × Recall Instruction interaction, F (10, 1062) = 10.57, $p < .001$. To summarize the immediate recall data, we can say that explicit F-item recall remains invariant over recall instruction, with the possible exception of the free-recall-all condition; A- and B-item recall

TABLE 1

PROBABILITY OF IMMEDIATE RECALL, FINAL RECALL, AND FINAL RECOGNITION AS A FUNCTION OF RECALL INSTRUCTION AND ITEM TYPE

Immediate recall instruction	Item type		
	F	A	B
Immediate recall			
A only	.040[a]	.292	.119[a]
B only	.031	.117[a]	.327
A and B	.038	.271	.302
A then B	.039	.271	.250
B then A	.033	.300	.283
Free recall all	.081	.275	.254
Final recall			
A only	.048	.121	.123
B only	.031	.123	.158
A and B	.051	.135	.125
A then B	.065	.144	.152
B then A	.042	.187	.131
Free recall all	.060	.100	.087
Final recognition[b]			
A only	.275	.435	.419
B only	.252	.369	.427
A and B	.275	.448	.406
A then B	.267	.448	.429
B then A	.248	.437	.437
Free recall all	.271	.391	.387

Note. Each point is based on 480 observations. See text for explanation of abbreviations.
[a] These figures represent intrusions, since *S*s were instructed not to recall the items in question.
[b] False alarm probability was .085.

does not. In the only conditions, *Ss* recalled less of the excluded items.

A 6 × 3 analysis of variance on the final recall data indicated a significant item effect, F (2, 1062) = 44.77, $p < .001$, but neither a significant effect of recall instruction, F (5, 1062) = 1.99, $p > .05$, nor a significant Item Type × Recall Instruction interaction, F (10, 1062) = 1.56, $p > .05$. The lack of interaction between item type and recall instruction indicates that the probability of recalling A, B, and F items remains invariant over recall instruction. The lack of an instruction effect suggests that the instruction effect obtained in immediate recall was short term and did not carry over into final recall. The significant item effect indicates that explicit F items were recalled less frequently than A or B items. When *Ss*, then, are asked to recall all items from the experiment, they still do not recall many explicit F items, although explicit F-item recall does increase slightly from immediate to final recall. This increase is entirely consistent with past studies using the final recall procedure (Woodward & Bjork, 1971). The B items from the only-A condition and A items from the only-B condition were recalled finally about as well as they were immediately.

The results of a 6 × 3 analysis of variance on the final recognition data were similar in nature to the analysis on the final recall data. A significant item effect was obtained, F (2, 1061) = 60.97, $p < .001$, but there was neither a significant effect of recall instructions, F (5, 1062) = .73, $p > .05$, nor a significant Item Type × Recall Instructions interaction, F (10, 1062) = .73, $p > .05$. Once again there were large differences between explicit and implicit F items, this time in their recognizability. Implicit F items were more likely to be recognized, just as they were more likely to be recalled. Final recall data, as well as final recognition data, are noteworthy, not in the small differences that occur for a given item type as a function of the recall instruction, but for the uniformity for a given item type across the various recall instructions. The F items clearly differ from A or B items in terms of their recallability and recognizability, but they do not differ as a function of the recall instruction. The A and B items do not differ as a function of recall instruction either, or from each other.

The data are not suggestive of an only effect. Epstein's (1969) only effect is defined as superior recall of the only condition compared with the appropriate both condition. Recall of A items in the only-A condition was .292, and in the A then B condition, .271. This difference was not significant, t (59) = .91, $p > .05$. Recall of B items in the only-B condition was not significantly greater than in the B then A condition, t (59) = 1.50, $p > .05$. Clearly, however, the items to be excluded at output are recalled, or actually intruded, far less frequently than items that are to be recalled. Thus, A-item recall in the only-B condition and B-item recall in the only-A condition is far less likely than in any other condition. The Item Type × Recall Instruction interaction stems largely from this low probability of recall of the items to be implicitly forgotten.

Finally, it should be noted that the data reported for the A then B condition and the B then A condition do not take into account order of recall. Although *Ss* were asked to recall (in accordance with the instruction) A items first followed by B items, or vice versa, recall of all items was scored. Taking order into account, and not accepting as correct A items intruded into B-item recall in the A then B condition, or B items intruded into A-item recall in the B then A condition, and the reverse, the probability of recalling A and B items was .183 and .214, respectively, in the A then B condition, and .210 and .194 in the B then A condition. There is some indication then that *Ss* had problems differentiating between A and B items.

Discussion

The fact that these data did not yield an only effect in retrospect is not surprising and is consistent with Epstein's research (1969, 1970). Epstein found that the only effect appears only when the two groups of items

are functionally separate. The interspersing of implicit items in this study through the use of the individual cuing technique resulted in a loss of the functional autonomy of each group of items and, subsequently, no only effect. Because no only effect was obtained, one could question the appropriateness of the term *implicit forget instruction.* In other words, is the presence of an only effect a necessary or prerequisite condition for the implicit forget instruction label? The present authors believe not. What is crucial are the operations required in producing the implicit forget instruction. In this regard, our operations were the same as those of Epstein's (1969). The instruction was given after input and before output, and in addition, *S*s did not know that they would ever be asked to recall or even recognize former list items. The recall-only-A instruction, as well as the recall-only-B instruction, although not successful in producing an only effect, were undeniably successful in reducing output of the excluded items.

The immediate recall data indicate that *S*s had very little difficulty separating implicit from explicit F items, but that making a similar discrimination within the implicit dimension (between A and B items) was considerably more difficult, although not impossible. Furthermore, final recall data indicate that the effect of the explicit instruction was long term, while the effect of the implicit instruction was short term. That is, recall of explicit F words was exceedingly low, both immediately and finally, while a low probability of recall of implicit F items immediately (i.e., a low recall of B items in the only-A condition and a low recall of A items in the only-B condition) did not reflect a similar trend in final recall.

The data strongly support the notion that explicit F items are different from implicit F items on indices of recall as well as recognition. Explicit F items are recalled rarely, immediately, or finally, and do not depend upon the recall instruction. Implicit F items are recalled with much higher probability than explicit F items, immediately and finally, and are also more likely to be recognized as old items. In addition, even when *S*s are asked immediately to recall explicitly labeled F items, they are less successful than when asked to immediately recall other types of items.

Epstein (1970) suggests that directed forgetting may be attributable to search and retrieval processes. Such an explanation may be appropriate for his pardigm which involves implicit postlist cuing and the only effect. However, the present authors would argue that the only effect is not representative of most directed forgetting experiments, and that implicit F items are not really forgotten. Directed forgetting, measured most reliably by either a final recall test or a special immediate recall test such as that which occurred with List 3, results with explicit within-list cuing, and its explanation involves rehearsal and other processing mechanisms at input, not search or retrieval processes at output.

The main difference, and a crucial one between explicit and implicit forget instructions, is the time of the instruction. In this experiment, explicit forget instructions followed the items immediately. The implicit forget instruction was delayed until the time of output and was far less effective in terms of leading to "forgetting." When the instruction is explicit, *S*s can stop further processing of the item. If the item is not processed, or processed minimally, it is not likely to be recalled or intruded even though it may be recognized. Items later followed by implicit forget instructions are processed by *S*s in the same way no doubt as items that they will attempt to recall. The *S*s, in an attempt to conform to the task requirements, exclude the implicit F items from recall, or attempt to, but this exclusion in no way affects later recall or recognition.

In conclusion, the experiment and the results reported in this paper point to clear differences both in recall and recognition between items that *S*s are explicitly told to forget as opposed to items that are implicitly to be forgotten. Put another way, there are vast paradigm differences among studies that fall under the label of intentional or directed forgetting and investigators should be sensitive to such differences.

REFERENCES

Bjork, R. A. Positive forgetting: The non-interference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 255–268.

Bjork, R. A., & Woodward, A. E., Jr. Directed forgetting of individual words in free recall. *Journal of Experimental Psychology*, 1973, 99, 22–27.

Bruce, D., & Papay, J. P. Primacy effect in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 473–486.

Davis, J. C., & Okada, R. Recognition and recall of positively forgotten items. *Journal of Experimental Psychology*, 1971, **89**, 181–186.

Elmes, D. G. Role of prior recalls and storage load in short-term memory. *Journal of Experimental Psychology*, 1969, **79**, 468–472.

Epstein, W. Poststimulus output specification and differential retrieval from short-term memory. *Journal of Experimental Psychology*, 1969, **82**, 168–174.

Epstein, W. Facilitation of retrieval resulting from postinput exclusion of part of the input. *Journal of Experimental Psychology*, 1970, **86**, 190–195.

Reitman, W., Malin, J. T., B[illegible], R. A., & Higman, B. Strategy control a[illegible]irected forgetting. *Journal of Verbal Lea[illegible] and Verbal Behavior*, 1973, **12**, 140–149.

Shebilske, W., Wilder, L., & Eps[illegible], W. Forget instructions: Effect of selective [illegible]hearsal and categorical distinctiveness. *Jour[illegible] of Experimental Psychology*, 1971, **89**, 372[illegible]

Woodward, A. E., Jr., & Bjork, R[illegible] Forgetting and remembering in free recall: I[illegible]ntional and unintentional. *Journal of Exp[illegible]ental Psychology*, 1971, **89**, 109–116.

(Received September 7, 1[illegible]

Journal of Experimental Psychology
1974, Vol. 10[illegible], No. 5, 1007–1014

EFFECTS OF IMAGERY INSTRUCTIONS, IMAGERY RATINGS, AND NUMBER OF DICTIONARY MEANINGS UPON RECOGNITION AND RECALL

M. J. PETERSON [1] AND S. H. McGEE

Indiana University

Three experiments investigated the effects of two "imagery-encouraging" variables, instructional set and imagery ratings, and of the number of dictionary meanings upon recognition and recall. Interactive imagery instructions produced superior recognition of stimulus words, response words, and pairs compared to rote-repetition instructions. Response recall also was higher following imagery instructions than following repetition instructions. High imagery ratings aided stimulus recognition and response recall, but not response recognition or pair recognition. Thus, the two imagery-encouraging variables had different effects upon recognition and recall and cannot be considered as two techniques for initiating the same underlying process. The effects of the number of dictionary meanings were independent of the effects of the two imagery-encouraging variables, although the meanings interacted with the imagery ratings of the pairs.

In recent years, two distinctly different techniques have been used to study "imagery" when the basic experimental materials were pairs of English nouns. One technique used nouns with scaled ratings of imagery (imagery ratings), and the other technique instructed *Ss* to imagine the referents of the nouns interacting in some way in their imaginary scene (imagery instructions). An interesting aspect of these two variables, which may or may not elicit imagery, is that they appear to have similar effects upon recall, despite the marked differences in experimental manipulations (e.g., Paivio & Yuille, 1969). The similarity of effects could be interpreted as the manifestation of a single underlying process that was induced by both variables, or as the reflections of independent but similar processes. Presumably, such identical or highly similar processes would influence various measures of retention, such as recognition and recall.

Suppose the variables trigger a single underlying process: both variables should then have similar effects on recognition and recall. Experiment I explored this possibility. As noted above, both variables apparently facilitate recall of the nouns compared to control conditions, but the effects upon recognition are less clear. Bower (1970), using only concrete nouns, reported no differences in recognition as a function of instructional set. However, his use of the concrete nouns may have produced a ceiling effect that precluded his finding a difference. The relevant recall literature suggested that words with low imagery ratings should be recognized less often than words with high imagery ratings when imagery ratings are varied, and these effects were verified by Peterson and Murray (1973). Their studies also showed better stimulus recognition following imagery instructions than following repetition instructions.

While the Peterson–Murray (1973) results indicated that both variables had similar effects upon recognition and recall, their experiments tested *stimulus* recognition (as did Bower's 1970 experiment). Experiment III of the present paper examined recognition of the response terms and of the pairs themselves as functions of imagery ratings and the instructional set.

Another purpose of the present studies was to probe the relationships between the two imagery-encouraging variables and a semantic characteristic of the words. We know that the imagery-encouraging variables are not reducible to the meaningful-

[1] Requests for reprints should be sent to M. J. Peterson, Department of Psychology, Indiana University, Bloomington, Indiana 47401.

ness or the familiarity of the words (see Paivio, 1971, for a review), or to lexical complexity (Kintsch, 1972). However, the imagery-encouraging variables may be related to a more cognitive index such as the conceptual complexity of the words. On an intuitive basis, words with high imagery ratings appear to differ in conceptual complexity from words with low imagery ratings, and the instructional set might produce similar differences. The possibility of relating the imagery variables to a single semantic index has theoretical appeal, of course. Consequently, Experiments II and III compared a measure of conceptual complexity, the number of dictionary meanings, with the two types of imagery-encouraging variables.

To summarize, the primary purposes of the series of studies were to investigate: (*a*) the effects of two imagery-encouraging variables upon the recognition of individual words and words in pairs and upon the recall of the word pairs, and (*b*) the symmetry of effects produced by the two imagery-encouraging variables and by the number of dictionary meanings.

Experiment I

Experiment I was designed to test recognition of the stimulus terms and recall of the pairs when the stimulus terms were shown alone, to provide an opportunity for imagining the referents of even low-imagery nouns, and when the pairs were then presented together. There were three phases. Phase 1 showed the stimulus items alone under instructions to imagine the referents of the words or to silently repeat the words. Phase 2 presented the stimulus items paired with the response words. Half of the *S*s given imagery instructions for Phase 1 were told to imagine the referents of the words interacting in their visual scene, and the other half were told to repeat the pairs silently. Similarly, *S*s who received repetition instructions before Phase 1 were split into imagery and repetition instruction groups for Phase 2. In Phase 3, all *S*s were tested for stimulus recognition and for response recall.

If imagery instructions and imagery ratings have essentially the same effect upon stimulus recognition [illegible] response recall, then both recogniti[illegible]nd recall should be higher with imager[illegible]structions and high imagery ratings th[illegible]ith rote-repetition instructions and [illegible] imagery ratings. No interaction woul[illegible]ear. On the other hand, if the va[illegible]s reflect independent operations, the[illegible]ht interact. For example, when *S*s re[illegible] imagery instructions, they may attem[illegible] imagine the referents (or some referen[illegible]gardless of the imagery ratings of the w[illegible]. When *S*s receive rote-repetition inst[illegible]ions, the words with high imagery ra[illegible] might induce the strategy of imaginin[illegible]e referents, while words with low ima[illegible] ratings would not. In this case, an [illegible]raction would be found such that the h[illegible]imagery items shown under imagery i[illegible]uctions would be recognized and recalle[illegible]ost accurately, and the low-imagery i[illegible] shown under rote-repetition instruction[illegible]uld be recognized and recalled least ac[illegible]ely.

Method

Design and materials. Four groups o[illegible] studied 36 nouns during Phase 1. Two of the[illegible]s were given instructions to visualize the obj[illegible]denoted by the words in a vivid way in their [illegible]l scene (imagery instructions). The two other [illegible]ps were told to repeat the words silently (ro[illegible]petition instructions). The groups then studied [illegible] pairs of nouns during Phase 2. One of the groups given imagery (I) instructions during Phase 1 and one of the groups given rote-repetition (R) i[illegible]uctions during Phase 1 received imagery instructions for Phase 2. These groups were identified as I–I and R–I, respectively. The two remaining groups received rote-repetition instructions just before Phase 2 began. These groups were designated as I–R and R–R.

The nouns shown in Phase 1 were the stimuli of the Phase 2 pairs. Thirty-six additional nouns were used as the Phase 2 response terms. Eighteen of the pairs were composed of concrete nouns, and the other 18 pairs contained abstract nouns. The concrete nouns had imagery ratings exceeding 6.00 and the abstract nouns had imagery ratings of less than 4.50, according to the Paivio, Yuille, and Madigan (1968) norms. The words were chosen to be approximately equal in rated meaningfulness. The high-*I* words were paired randomly to yield 18 pairs, and the low-*I* pairs were made up in the same way. Eighteen additional high-*I* words and 18 more low-*I* words were used as distractors in the stimulus recognition test. These words met the criteria stated above.

The words were typed on mats for slide presentation. The mats were ordered randomly for Phases

TABLE 1

...rmance on Stimulus Recognition and Response Recall Tests: Experiment I

...ery ratings	Stimulus recognition			Response recall	
	Mean hits	Mean false alarms	A'	Mean correct	Conditional correct
...oup I–I					
High	17.19	.44	.98	8.50	.49
Low	17.44	.38	.98	5.88	.34
...roup I–R					
High	17.19	.94	.97	5.28	.31
Low	16.75	1.25	.96	2.31	.14
Group R–I					
High	16.00	.88	.95	6.69	.41
Low	16.06	2.56	.90	3.12	.20
Group R–R					
High	15.38	1.88	.94	3.50	.21
Low	15.81	3.31	.91	1.44	.09

a For group names, the first letter indicates Phase 1 instructions, the second Phase 2 instructions. I = imagery, R = rote ...ition.

1 and 2 and for the stimulus-recognition and response-recall tests. All *Ss* received the same ordering of mats for each phase.

Procedure. The *Ss* were told that there were three phases in the experiment and that they should remember the items of each phase. They received the appropriate instructions and were shown the slides of Phase 1 at a 5-sec. rate. Following Phase 1, *Ss* were read their assigned instructions for Phase 2, and the Phase 2 slides were shown at an 8-sec. rate. After a 60-sec. break, instructions were given for the recognition-recall tests. The *Ss* were told to write down the stimulus word as it was shown on the screen. They then indicated their certainty that the word had been shown during Phases 1 or 2 by hecking a 5-point scale: Certainly Old, Probably ld, Don't Know, Probably New, and Certainly ew. The next task was to recall the word paired ith the word written on the sheet, if *Ss* thought hey had seen the stimulus word previously. Each lide was shown for 15 sec. during the recognition-ecall task.

Subjects. Students in introductory psychology courses at Indiana University participated as one method of fulfilling a course requirement. Sixty-four *Ss* were randomly assigned to 16 replications of the block of four groups.

Results

Stimulus recognition. The mean numbers of correct stimulus recognitions (hits), the mean numbers of incorrectly labeled new stimuli (false alarms), and A' estimates[2] (Pollack & Norman, 1964) are given in Table 1. The hit and false-alarm rates were calculated from judgments of Certainly Old and Probably Old given to true old stimuli and to the distractors, respectively. The mean number of hits for the groups receiving imagery instructions before Phase 1 reliably exceeded the mean number of hits for the groups with rote-repetition instructions for Phase 1, $F\ (1, 60) = 214.28$, but neither the instructions given for Phase 2, $F < 1$, nor the imagery ratings, $F\ (1, 60) = 1.02$, $p > .05$, produced significant differences in the hit rates.

The false alarms varied as a function of the instructions given for both phases. The false-alarm rates were higher following rote-repetition instructions given before Phase 1, $F\ (1, 60) = 55.86$, $p < .01$, and before Phase 2, $F\ (1, 60) = 16.93$. The imagery ratings did not have a reliable effect upon the false-alarm rates.

The more frequent labeling of a new stimulus as old following rote-repetition instructions suggested that *Ss* in these groups were using a more lax criterion for judgment than *Ss* in the imagery-instruction groups. To explore detection sensitivity further, A' estimates were obtained for each condition, and these estimates were subjected to an analysis of variance.

The mean A' estimate for groups receiving imagery instructions before Phase 1 was significantly greater than the mean A' for groups with rote-repetition instructions before Phase 1, $F\ (1, 60) = 34.00$, $p < .01$. The instructions that preceded Phase 2 did not show significantly different estimates of A', $F\ (1, 60) = 2.00$, $p > .05$, nor was

[2] The authors wish to thank Donald E. Robinson for supplying the tables of A'.

the interaction between the Phase 1 and Phase 2 instructions reliable ($F < 1$).

The other method of inducing imagery, the imagery ratings, also showed significant effects upon stimulus recognition when the bias toward labeling new stimuli as old was taken into account. The A' estimate was reliably higher for stimuli with high imagery ratings than for stimuli with low imagery ratings, $F\ (1, 60) = 20.62, p < .01$. Further, the imagery ratings interacted with the Phase 1 instructions, $F\ (1, 60) = 12.00, p < .01$. The mean A' scores for the conditions contributing to this interaction were high-*I* words, imagery instructions (.98), low-*I* words, imagery instructions (.97), high-*I* words, repetition instructions (.95), and low-*I* words, repetition instructions (.91). Tukey tests indicated that the means for high- and low-*I* words under imagery instructions did not differ reliably: all other comparisons were significant at the .01 level. These results indicated that when *S*s received imagery instructions, recognition performance was uniformly high, regardless of the imagery ratings of the pairs. The imagery ratings of the pairs became an important determinant of recognition performance only under rote-repetition instructions. No other interactions were reliable.

Response recall. Table 1 contains the mean numbers of correctly recalled responses and the proportions of correctly recalled responses conditionalized upon correct stimulus recognition. Analyses showed the same effects for both response measures, hence, only the second measure will be discussed. Reliably more responses were recalled following correct stimulus recognition under imagery instructions than under rote-repetition instructions for both Phase 1 and Phase 2: for Phase 1, $F\ (1, 60) = 40.00, p < .01$; for Phase 2, $F\ (1, 60) = 152.24, p < .01$; and when the pairs had high imagery ratings, rather than low imagery ratings, $F\ (1, 60) = 14.70, p < .01$. No interactions were reliable.

Discussion

As expected, both imagery-encouraging variables were associated with increased response recall, and the two variables did not interact. These results imply that t[illegible] variables reflect the same or highly s[illegible]processes, but the stimulus recognition d[illegible]ided contradictory evidence. Phase 1 i[illegible]y instructions facilitated stimulus recogni[illegible]re than rote-repetition instructions, but [illegible]gery ratings produced differences only wh[illegible]mbined with rote-repetition instructions. [illegible]nce, these two variables have different effe[illegible]pon stimulus recognition and must be a[illegible]ed to induce distinct underlying process[illegible]

Phase 2 instructions had [illegible] differential effects upon stimulus recogni[illegible]. The most plausible explanation for the [illegible]e of Phase 2 instructions to influence sti[illegible] recognition was that *S*s who received ima[illegible]nstructions before Phase 1 were likely t[illegible]inue to use an imagery strategy during Ph[illegible], regardless of the instructions that imme[illegible]ly preceded Phase 2. The carry-over fr[illegible]ne instructional set to the other may be [illegible]ated from consideration by comparing [illegible]ition performance on high-*I* and lo[illegible] words for Groups I-I and R-R. This co[illegible]son showed the same effects as those desc[illegible] above for stimulus recognition (see Tab[illegible]).

Experiment I may have und[illegible]imated recall, particularly for the groups [illegible]ving rote-repetition instructions. It is [illegible]ible that when *S*s were uncertain of their [illegible]gnition of the stimulus words or were unable to recognize the stimulus words, they did not try to remember the responses, thus suppressing recall. Experiments II and III were designed to avoid these potentially confounding effects.

Experiment II

Experiment II asked if the imagery-encouraging variables could be subsumed by the number of dictionary meanings of the words of the pair. Such a prediction would oppose the contemporary theories (e.g., Atwood, 1971; Bower, 1972; Paivio, 1971), which assume visual imagery and verbal attributes to be distinct but interacting processes. If these processes do interact extensively, then questions about the likelihood of enhanced encoding or whatever occurs when the imagery-encouraging variables are manipulated might better focus upon the conjoint relationship between measures of the verbal–linguistic characteristics of the words and of imagery ratings.

It is possible that the more meanings a word has, the greater are the chances that

so… of these characteristics will be imaginable. When an easily imagined noun has … eanings, or when *S*s attempt to imag… the referents of a noun with many …ings, these meanings may "enrich" … image or they may produce multiple images. Other easily imagined words with fewer meanings (or less meaningful words whose referents are to be imagined) would be expected to yield fewer images or images with less enrichment. On the other hand, the possession of many meanings might interfere with the encoding and acquisition of words which are more difficult to imagine.

Thus, Experiment II manipulated the number of dictionary meanings, instructional set, and the imagery ratings of the pairs of words. The ratings of the *pairs* were assumed to provide a better index of the saliency of an image combining the referents of the two words than an average of the imagery ratings of the individual words of the pair. The imagery ratings of the individual words were held relatively constant, while the rating of the pair varied. For example, JUDGE–TABLE had a high imagery rating as a pair and JUDGE–WHEAT had a low rating as a pair, although the imagery ratings of the individual words were approximately the same.

Method

Preliminary scaling and design. Eighty nouns with imagery ratings of 6.00 or higher and with meaningfulness ratings of 5.50 or higher were drawn from the Paivio et al. (1968) norms. Seventy-two randomly selected "stimulus" nouns were then paired with each of the eight remaining "response" nouns. The imagery ratings of these pairs were obtained by administering booklets containing the pairs to 30 students in introductory psychology courses at Indiana University. Six different booklets were used so that a particular stimulus noun would be paired with only one of its response words per booklet. The instructions for rating were similar to those used by Paivio et al., and *S*s rated the imagery of the pairs on a 7-point scale.

In addition, the numbers of different "meanings" or separate entries found in *Webster's Unabridged Dictionary* (1966) were tabulated. This variable was labeled "dictionary meaning" (*dm*). The numbers of dictionary meanings were calculated for pairs of words by adding the number of dictionary meanings for the two words of each pair. Finally, 12 pairs were selected for each of the four cells of the design that arose from the two levels of imagery ratings (high and low) and the two levels of dictionary meanings (high and low). The mean values are given in Table 2 for each cell. Table 2 also shows sample pairs. It should be noted that for the imagery variables, each stimulus noun was paired with two responses. This technique was used to control the imagery ratings of the stimuli and because Experiment III was to test recognition of the response words.

TABLE 2

MEAN IMAGERY RATINGS, MEAN NUMBER OF DICTIONARY MEANINGS, AND SAMPLE PAIRS OF MATERIALS: EXPERIMENT II

Condition	Imagery ratings[a]	Dictionary meanings	Sample pair
High *I*–high *dm*	5.76	13.62	JUDGE–TABLE
High *I*–low *dm*	5.74	6.92	DOCTOR–PUPIL
Low *I*–high *dm*	2.70	13.04	DOCTOR–EARTH
Low *I*–low *dm*	3.20	6.96	JUDGE–WHEAT

Note. Abbreviations: *I* = rated imagery, *dm* = number of different dictionary meanings.
[a] Maximum score for high imagery was 7.00.

Two lists were prepared, using the first pairing for a stimulus in one list and the other pairing in the second list. Each list contained six high-*I*, high-*dm* pairs; six high-*I*, low-*dm* pairs; six low-*I*, high-*dm* pairs; and six low-*I*, low-*dm* pairs.

Subjects. Sixty-four new *S*s from the same pool were assigned randomly to 16 replications of the block of two lists and two types of instructions.

Procedure. The *S*s were given either imagery or rote-repetition instructions. They were then shown slides with the noun pairs at the rate of 8-sec. per slide. After about a 2-min. interval, the stimuli were presented alone, and *S*s were instructed to write the appropriate response on their answer sheets. The procedure of Experiment II was comparable to Phase 2 of Experiment I, followed by a test for response recall. Experiment II had no Phase 1 counterpart, nor was any test for stimulus recognition introduced.

Results

The major results are shown in Table 3. The *S*s given imagery instructions recalled significantly more responses than *S*s given rote-repetition instructions, F (1, 60) = 34.17, $p < .01$, and more responses were

TABLE 3

MEAN NUMBER OF CORRECTLY RECALLED RESPONSES: EXPERIMENT II

Dictionary meanings	Imagery instructions		Repetition instructions	
	High *I*	Low *I*	High *I*	Low *I*
High	4.81	4.03	3.22	1.97
Low	4.50	4.38	2.84	2.59

Note. Abbreviation: *I* = rated imagery.

recalled from the high-I pairs than from the low-I pairs, $F(1, 60) = 20.47$, $p < .01$. The latter result indicates that response recall reflects the imagery ratings of the pairs, not simply the ratings of the individual words of the pairs. No other main effect was reliable.

The failure to demonstrate a significant main effect for the number of dictionary meanings indicated that this variable did not explain the effects of the two imagery-encouraging variables upon response recall. However, the number of dictionary meanings did interact with the imagery ratings of the pairs, $F(1, 60) = 12.42$, $p < .01$. The high-I pairs were recalled more often when they had more dictionary meanings than when they had fewer dictionary meanings, but the opposite was true for low-I pairs, as shown in Table 3. The number of dictionary meanings did not interact with the instructional set.

Discussion

The results of Experiment II supported theories which posit independent visual and verbal components or processes and argued against the possibility that the imagery-encouraging variables might be reduced to a measure of conceptual complexity such as the number of dictionary meanings. The interaction between the imagery ratings of the pairs and the number of dictionary meanings may have been produced by increases in the number of meanings which "enriched" the image or produced multiple images, while the extra meanings hindered the processing of words which were difficult to imagine.

Experiment II also showed that the imagery ratings of the pairs of words predicted recall more satisfactorily than the imagery ratings of the individual words, providing additional support for the notion that the association of the words is important. Experiment III investigated recognition of response terms and of the pairs under the same conditions as Experiment II.

Experiment III

Experiment III presented pairs of words with high and low numbers of dictionary meanings and with high and low ratings of imagery. During the exposure trials, one group received imagery instructions and the other group, repetition instructions. Half of each group was given a response recognition test and half, a pair recognition test. If the imagery-encouraging variables affect the encoding of the words, subsequent recognition of the individual words and of the pairs should be better for pairs shown under imagery instructions or for pairs with high imagery ratings compared with pairs shown under rote-repetition instructions or for pairs with low imagery ratings. These predictions have been tested for the *stimulus* items (Bower, 1970; Peterson & Murray, 1973; Experiment I of the present series). Experiment III examined the effects upon recognition of the response terms and of the pairs.

The number of dictionary meanings was included because other investigators have reported that recognition was inversely related to the frequency of occurrence of words (Shepard, 1967) and to the meaningfulness of dissyllables (McNulty, 1965), while recall showed a positive relationship. The finding that dictionary meanings mirrored these results while imagery-encouraging variables did not would be additional evidence that the imagery-encouraging variables differ from the various measures of verbal–linguistic characteristics.

Method

Materials and design. The materials were those used in Experiment II. Half of the *Ss* saw List 1 during the acquisition stage and the other half saw List 2. All *Ss* were then transferred to a recognition list that contained all of the pairs or responses from Lists 1 and 2. Thus, on the recognition test, half of the items would be old and half would be new, for all *Ss*.

One group was told to imagine the referents of the words, and a second group was told to repeat the words silently. These two groups were subdivided for the recognition test. Half of each group received the response recognition test and half, the pair recognition test.

Subjects. One hundred and ninety-two new students from the same pool were assigned to 24 replications of the block of eight cells produced by the two lists and the four instruction groups.

Procedure. After hearing the appropriate instructions, the pairs of the acquisition list were presented at an 8-sec. rate. The *Ss* were reminded to use imagery or rote repetition during the recognition test to help them make their decisions. The *Ss* rated their confidence of their judgments by

TABLE 4

PERFORMANCE ON RESPONSE AND PAIR RECOGNITION TESTS: EXPERIMENT III

Performance measure	Imagery instructions				Repetition instructions			
	High *I*		Low *I*		High *I*		Low *I*	
	High *dm*	Low *dm*	High *dm*	Low *dm*	High *dm*	Low *dm*	High *dm*	Low *dm*
Response recognition								
Mean hits	5.38	5.46	5.25	5.50	4.75	4.58	4.21	4.75
Mean false alarms	2.04	.79	1.50	1.00	1.96	.96	1.62	1.38
A′	.84	.94	.88	.93	.80	.85	.78	.85
Pair recognition								
Mean hits	5.71	5.91	5.71	5.71	5.21	5.08	5.42	5.21
Mean false alarms	.08	.08	.21	.16	.29	.25	.33	.33
A′	.98	.99	.97	.98	.95	.94	.95	.94

Note. Abbreviations: I = rated imagery, dm = number of different dictionary meanings.

circling numbers on the same 5-point scale used in Experiment I. No measure of response recall was obtained in Experiment III.

Results and Discussion

The mean numbers of hits, false alarms, and the A' estimates are given in Table 4 for the four groups. The recognition of the response words was significantly better following imagery instructions than following repetition instructions, F (1, 46) = 13.64, $p < .01$, and for words with *few* rather than many dictionary meanings, F (1, 46) = 34.16, $p < .01$. Imagery ratings did not have a differential effect upon response recognition ($F < 1$). No interactions were significant.

The finding that words with few dictionary meanings were recognized with greater accuracy than words with many dictionary meanings, while the opposite was true for recall, corresponded to similar results reported by Schwartz and Rouse (1961) and Shepard (1967) for the frequency of occurrence of words and by McNulty (1965) for the meaningfulness of dissyllables.

In general, detection of old from new pairs was superior to the detection of old from new responses. Imagery instructions also facilitated the recognition of the pairs, F (1, 46) = 5.95, $p < .05$, but neither the imagery ratings, F (1, 46) = 1.00, $p > .05$, nor the dictionary meanings, $F < 1$, were effective in enhancing the detectability of the pairs. These results make it clear that the response recall found in Experiment II for these pairs could not be explained by simple recognition of either the response term alone or of the pair as a unit.

General Discussion

The results of the three experiments indicated that theoretical explanations of imagery must accord separate treatments to the two imagery-encouraging variables, instructional set and imagery ratings, because the two variables had different effects upon recognition. Their effects upon recall were highly similar and did not interact.

Imagery instructions aided recognition of the stimulus words, the response words, and the pairs to a greater extent than control (rote-repetition) instructions. Imagery ratings were directly related to recognition of the stimulus words experienced under rote-repetition instructions but not to recognition of the stimulus words shown under imagery instructions or to recognition of the responses or the pairs, irrespective of the instructional set. Clearly, the instructional variable has a wider scope of influence than the imagery ratings, and the two techniques for encouraging imagery induce processes which have differential effects upon recognition.

Despite their disparate effects upon recognition, the imagery-encouraging variables showed a highly similar influence on recall. Significantly more responses were correctly recalled following imagery instructions than following repetition instructions and for pairs with high rather than low imagery ratings.

Why should the two imagery-encouraging variables have different but interacting effects upon recognition but not upon recall? One

explanation is that one of these variables, imagery instructions, influences both encoding and retrieval, thus affecting recognition and recall, while the other variable, imagery ratings, exerts its primary effect upon retrieval. This explanation would have to be modified to permit imagery ratings to affect encoding when no imagery instructions were in effect (see Paivio, 1971, for a review).

A varient hypothesis is that imagery instructions induce a process that facilitates both recognition and recall. Imagery ratings, of secondary importance, are effective only to the extent that they induce the same process. This hypothesis predicts that recall performance should mimic recognition performance, in contradiction to the present results. A slightly different version of this hypothesis is that recall of the pairs is facilitated by both variables, through enhanced recognition of the cue term, and that imagery instructions are dominant over imagery ratings. Further, imagery instructions simply have additional effects, such as the enhancement of the responses and of the pairs, that are irrelevant to recall. This possibility seems most reasonable.

The process or processes that underlie the imagery-encouraging variables may or may not involve visual components. It is quite plausible that the imagery instructions, for example, trigger active, but not necessarily visual, rehearsal mechanisms on the part of *S*, whereas the rote-repetition instructions elicit a far more passive process. So long as the process produces sufficient encoding, recall will be possible. The more active rehearsal might well yield differential confidence ratings of recognition performance, but recall would not differentiate between rehearsal and more passive repetition if the items were encoded at some minimal level.

Neither imagery-encouraging variable was explainable by a measure of conceptual complexity such as the number of dictionary meanings. Increases in the number of dictionary meanings facilitated recall but hindered recognition of the responses, whereas the imagery instructions, in particular, were positively related to both recall and recognition of the responses. These results also refuted the earlier proposal that high-imagery words with many meanings might yield more encodings or an enriched encoding. Instead, the encoding accorded high-imagery words with many meanings presumably is less distinctive than the encoding of high-imagery words with a limited number of dictionar[illegible]ings. To explain the superior recall [illegible]e words despite the less accurate recogni[illegible] is necessary to assume that retrieval [illegible]gies are more effective with the high-[illegible], high-meaning words.

Thus, we are faced with [illegible]triguing dilemma of acknowledging [illegible]lapping, sometimes contradictory effects [illegible] recognition and recall of two variables [illegible]legedly induce imagery processes witho[illegible]ng able to assess precisely why these [illegible]bles are effective and without being able [illegible]ie these variables to other characteristic[illegible] verbal materials known to affect recog[illegible]on and recall.

REFERENCES

ATWOOD, G. An experimental stud[illegible] visual imagination and memory. *Cognitive [illegible]hology*, 1971, 2, 290–299.

BOWER, G. H. Imagery as a relational o[illegible]izer in associative learning. *Journal of Verb[illegible]arning and Verbal Behavior*, 1970, 9, 529–533.

BOWER, G. H. Mental imagery and [illegible]iative learning. In L. Gregg (Ed.), *Cognitio[illegible] Learning and memory.* New York: Wiley, 19[illegible]

KINTSCH, W. Abstract nouns: Imag[illegible]ersus lexical complexity. *Journal of Verbal [illegible]rning and Verbal Behavior*, 1972, 11, 59–65.

McNULTY, J. A. An analysis of recall and recognition processes in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 1965, 4, [illegible]435.

PAIVIO, A. *Imagery and verbal proces[illegible]* New York: Holt, Rinehart and Winston, 1971

PAIVIO, A., & YUILLE, J. C. Changes in as[illegible]ative strategies and paired-associate learning ov[illegible] trials as a function of word imagery and type of learning set. *Journal of Experimental Psychology*, 1969, 79, 458–463.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, 76(1, Pt. 2).

PETERSON, M. J., & MURRAY, A. Enhancement of items and associations. *Journal of Experimental Psychology*, 1973, 98, 64–69.

POLLACK, I., & NORMAN, D. A. A non-parametric analysis of recognition experiments. *Psychonomic Science*, 1964, 1, 125–126.

SCHWARTZ, F., & ROUSE, R. D. The activation and recovery of associations. *Psychological Issues*, 1961, 3 (Whole No. 1.)

SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 156–163.

Webster's third new international dictionary of the English language, unabridged. Springfield, Mass.: G. & C. Merriam, 1966.

(Received September 10, 1973)

Journal of [illegible] *Psychology*
1974, Vol. 1[illegible], 1015–1020

[illegible]ARALLEL PROCESSING IN A WORD–NONWORD CLASSIFICATION TASK[1]

NEIL NOVIK[2]

University of Connecticut

Undergraduate *S*s were required to indicate whether each stimulus in a series of CCCs, CVCs, WORDs (CVC type) and meaningful CCCs (e.g., JFK, LSD) was a word or a nonword. Reaction time (RT) was the dependent measure. The serial model of word–nonword classification, which states that phonological evaluation precedes memory search, would predict no differences in RT between the meaningless CCCs and the meaningful CCCs, since the two items are equally unlawful. The parallel model, which assumes that phonological evaluation and memory search occur simultaneously, would predict that RTs to meaningful CCCs would be greater than those to meaningless CCCs, because for the former type of item, there is confusion in finding meaning in an unlawful nonword. The results supported the parallel model in that RTs to meaningful CCCs were consistently greater than those to meaningless CCCs. In a second experiment, RTs to frequent CCCs equalled RTs to nonfrequent CCCs, suggesting that the word nonword decision was not based on a frequency determination.

A considerable amount of attention has recently been paid to what Norman (1969, pp. 162–163) has termed the "mantiness problem." A person is able to recognize in a very short amount of time that "mantiness" is not a word in his vocabulary. The problem, of course, is how this accurate decision is arrived at so quickly. It has therefore been the goal of several recent studies to discover the processes underlying this ability.

When confronted with "mantiness" it is highly unlikely that the person engages in a scan of all the words he has stored in memory (this process would take too much time), but it is generally agreed that recognition of a word or rejection of a phonologically lawful nonword requires some search for meaning. In addition, several authors (Norman, 1969; Rubenstein, Lewis, & Rubenstein, 1971; Stanners, Forbach, & Headley, 1971) suggest that rejection of a phonologically unlawful item such as "mgsptzy" may be accomplished by means of some preprocessing of the item which would determine its phonological lawfulness.

The present paper deals with the question of how this preprocessing and the memory search may be related. The two basic alternatives are that the operations are sequential or that they act in parallel. In the sequential model, the person first decides whether the item could be a word, according to phonological rules. If a stimulus is rejected, no search through memory for meaning is needed. If, on the other hand, the item is accepted as being phonologically lawful, the person must search for meaning. A *word* response would be emitted if meaning was found, while a *nonword* response would follow the failure to find such meaning. This type of model has been described by Rubenstein et al. (1971) and Stanners et al. (1971). Its support comes mainly from studies in which *S* is required to respond to a visually presented stimulus simply on the basis of whether the item was a word or a nonword. The model predicts that reaction times (RTs) for rejecting phonologically unlawful units (e.g., consonant–consonant–consonants—CCCs) would be faster than for accepting words, since no search for meaning is required in the former case. Furthermore, RTs to

[1] This research was supported in part by National Institutes of Health Training Grant 00084. Part of this paper was presented at the meeting of the Eastern Psychological Association, Washington, D. C., May 3–5, 1973. The author wishes to express his gratitude to Ron Lecours for his assistance in the collection of data and to his advisory committee for their helpful suggestions.

[2] Requests for reprints should be sent to Neil Novik, Department of Psychology, University of Connecticut, Storrs, Connecticut 06268.

phonologically lawful nonwords (e.g., consonant–vowel–consonants—CVCs) should be slower than those to words, since an exhaustive search is needed when the item is not a word.

These predictions have been supported in several recent studies. Lawful nonwords have been shown to have longer RTs than both real words (Rubenstein, Garfield, & Millikan, 1970; Snodgrass & Jarvella, 1972) and unlawful nonwords (Rubenstein et al., 1971). The strongest support for the sequential model comes from the study by Stanners et al. (1971), in which *S*s responded *word* or *nonword* to a series of visually presented CCCs, words (CVC type) and CVCs (nonwords). As the sequential model predicts, CCCs were responded to the fastest, words were intermediate, and CVCs were the slowest (approximate RTs were 600, 715, and 840 msec., respectively). The authors also found that the frequency in the English language of the initial and terminal consonants of the stimuli affected the RTs for words and CVCs but not for CCCs. These results suggested that phonological evaluation precedes the search for meaning and that the former operation is unaffected by at least one variable (consonant frequency) affecting the latter operation.

The parallel model assumes that phonological information and the search for meaning are processed simultaneously. Although, at present, there is no direct support for such a model, several lines of evidence have demonstrated that familiarity can have a very early effect in the processing of visual information. At least two studies have shown that the recognition of individual letters is improved when those letters are embedded in words (Reicher, 1969; Wheeler, 1970). Words have also been shown to have an advantage over nonwords in physical matching (Eichelman, 1970) and visual scanning (Krueger, 1970; Novik & Katz, 1971).

These studies are important in showing that higher order information such as meaning may have very early effects in visual information processing. Although not directly supporting the parallel model described earlier, these data [illegible] st that it might be possible for the [illegible] ing of a given item to be processed ea[illegible] recognition, even simultaneously w[illegible] onological information.

The present study seeks to [illegible] le direct support for that type of p[illegible] model. Specifically, it was hypothes[illegible] hat two kinds of information, one co[illegible] g phonological lawfulness and the [illegible] her concerning the meaning of the i[illegible] may be processed simultaneously. Th[illegible] riment is based on the procedure used [illegible] anners et al. (1971). The *S*s were pr[illegible] l with a series of three-letter items [illegible] ad to indicate whether each item wa[illegible] rd or a nonword. The dependent [illegible] e was RT. As in the Stanners et al [illegible] er, *S*s were presented with CCCs, [illegible] and CVC words. In contrast to Stan[illegible] et al., *S*s in the present study also sa[illegible] group of CCCs which presumably co[illegible] ed a good deal of meaning (e.g., J[illegible] LSD, etc.). The main comparison in th[illegible] esent study was between the RTs to t[illegible] meaningful CCCs and those to the non[illegible] ningful CCCs. If a determination o[illegible] honological lawfulness were made prio[illegible] o the search for meaning, and *S* was able to reject an item on the basis of that determination, then the RTs to JFK and JRK should not differ. This is true because the items are presumably equally unlawful and because meaning does not affect the phonological decision in the sequential model. If, however, the two operations occurred in parallel, then the finding of meaning for JFK would conflict with the decision of phonological unlawfulness, causing a delay in the nonword response. In summary, the sequential model would be supported by the finding of no differences in the RTs to the two types of CCCs, while such a difference would tend to support the parallel model.

Experiment I

Method

Subjects. Twenty right-handed female undergraduate students enrolled in introductory psychology at the University of Connecticut served as *S*s in the experiment. Each *S* received 1 hr. of experimental credit toward the course requirement.

Apparatus. The main piece of apparatus was a Kodak Carousel slide projector with a Pronto Press tachistoscopic shutter. The equipment used in programming the slide presentations consisted of two Lafayette 100-sec. timers, a Lafayette triple pulse timer, and a Hunter interval timer.

There were three telegraph keys placed on a table in front of *S*. The center key was used to initiate a trial, while the two outside keys were used for making responses. A Hunter digital clock measured in milliseconds the interval between the presentation of the stimulus and the response.

The slides were projected onto a rear-projection screen placed 5 ft. from *S*. The screen measured 14 × 14 in., but all except a center portion of the screen, measuring 5 × 3 in., was blocked out by cardboard taped to the back. The projected image of the stimulus was 3 × 1 in. and was projected in the center of the screen.

[illegible] The stimuli consisted of WORDs (CVC type), CCCs, and CVCs (nonwords), typed in uppercase letters and copied onto transparency film by a 3M Thermofax copier. No stimulus contained more than one of the same letters.

There were two 60-trial practice blocks. In each block there were 30 WORDs, 15 CCCs, and 15 CVCs. The CCCs and CVCs were constructed by randomly selecting 15 of the WORDs and inserting a random consonant or vowel in the center position (the only restriction being, of course, that for the CVCs, another WORD did not result). For example, if CAN was the WORD, the CCC and CVC could have been CHN and CEN, respectively. Thus, in each practice session, 15 of the WORDs had two corresponding nonwords, while the other 15 WORDs were filler items.

For the test trials, 15 meaningful CCCs (e.g., JFK, LSD) were chosen. The selection of these items was based on the results of questionnaires in which pilot *Ss* were asked to indicate the meaning of 20 such meaningful CCCs. The five items responded to most incorrectly were eliminated. From each meaningful CCC (CCC+) another CCC and a CVC were constructed in the same manner as described for the practice stimuli. In addition, 45 WORDs were chosen for the test trials independently of the nonwords, since several of the CCC+s could not be transformed into words (e.g., JFK, NBC, LBJ). Thus, there were 90 test trials: 15 CCC+s, 15 CCCs, 15 CVCs, and 45 WORDs. Within each practice block and the test block the stimuli were randomly mixed. Each *S* received the same order of stimuli.

Procedure. The *S* was instructed that pressing the center key would initiate the trial, and that 2 sec. after she pressed the key, the stimulus would be presented. The *Ss* were told to respond to that stimulus as quickly and as accurately as possible on the basis of whether it was a word or a nonword. The key on the right was designated for a *word* response, while the left key was to be used for a *nonword* response. The *S* used her right hand for a *word* response and her left hand for a *nonword* response. There were no requirements as to which hand was to be used to press the center key, but *S* was instructed that both hands should be ready to respond when the slide was presented. No *S* reported any difficulty in getting ready for each presentation.

During the first practice block, speed and accuracy were emphasized by *E* approximately every 15 trials. There was a 5-min. rest period between the two practice blocks. There was also a 10-min. interval between the second practice block and the test trials, during which time *Ss* were told that for the test trials, they would receive 4¢ for every RT under 900 msec., and that 25¢ would be taken away for every error. There was a slight delay midway through the test trials, long enough for *E* to change slide trays.

For each trial, a maximum of 4 sec. was allowed for a response. The failure criterion for the test trials was 10 errors. Two *Ss* made more than 10 errors and were replaced.

TABLE 1

Mean Reaction Times (RTs) (in msec.), Standard Deviations, and Percent Errors for the Test Trials in Experiment I

Stimulus type	Number of stimuli	Mean RT	*SD*	% error
CCC+	15	695	66	1.6
CCC	15	672	59	1.6
WORD	45	677	47	3.9
CVC	15	785	78	11.6

Note. Abbreviations: C = consonant, V = vowel, + = meaningful.

Results

Median RTs of correct responses for the test trials were obtained for each *S* under each condition. Table 1 presents the RT data as well as the error data for the test trials. A one-way repeated measures analysis of variance was performed on the RTs to the three nonword conditions. The RTs to the WORDs were eliminated from the analysis because all of the WORDs were filler items, i.e., they were chosen independently of the nonwords. The analysis revealed a highly significant treatment effect, $F(2, 38) = 82.67$, $p < .001$. Comparisons between conditions were made by a series of Duncan's multiple range tests. As in several of the studies previously cited, the RTs to the lawful nonwords (CVCs) were longer than those to the unlawful nonwords (CCCs and CCC + s) ($p < .01$). More

importantly, however, the RTs to the CCC + s were significantly longer than those to the CCCs ($p < .05$). The latter effect was also found in individual trends, with 18 of the 20 *S*s having a greater RT to the CCC + s than to the CCCs. The CCC+ vs. CCC difference was also found with individual pairs of stimuli. Median RTs across *S*s for each CCC+ and CCC were obtained, and of the 15 pairs of stimuli, the RTs to 12 CCC + s were greater than those to their corresponding CCCs, while 3 pairs showed the reverse trend. Both of these proportions are significant ($p < .025$) according to sign tests.

In order to compare the RT data for WORDs to the data for the other conditions, a series of related *t* tests were performed on the individual medians. The RTs to WORDs were found to be significantly shorter than those to CVCs, $t(19) = 8.70$, $p < .001$, but no other differences were found.

A series of sign tests showed that there were significantly more errors made with the CVCs than with the other stimuli ($p < .001$), while no other differences in percent errors were found. Since the RTs of incorrect responses might be indicative of strategies in processing (Stanners et al., 1971), median RTs for the incorrect responses were obtained for the CVCs and WORDs. The medians were based on a total across all *S*s of 35 incorrect responses in both conditions. For the CVC errors, the median was 711 msec., which was considerably lower than the mean RT for CVC correct responses. The median for WORD errors, on the other hand, was 756 msec., which was greater than the mean RT for correct WORD responses. For both the CCC and CCC+ conditions, there were too few errors for a median to be meaningful.

Discussion

The most important finding of the experiment was that the RTs to the meaningful, but unlawful, items (CCC + s) were greater than those to the unlawful meaningless items (CCCs), a difference that would not be predicted by the sequential model described above. According to that model, there is a two-stage process in word–n[illegible]l classification. During the first stag[illegible]honological structure of the item is eval[illegible] and if the item is found to be phonolog[illegible] unlawful, a nonword response may be e[illegible]ed. If the item is found to be lawful, *S* s[illegible] his memory in order to determine if the i[illegible] is indeed a word.

In the present experiment, th[illegible]C + s and the CCCs were presumably e[illegible] unlawful, and therefore, according to suc[illegible]odel, both types of items should have been [illegible]ted in the first stage, yielding equal RT[illegible]e finding of a consistent difference in R[illegible]tween the CCC+ and CCC conditions brin[illegible]e sequential model into serious doubt. [illegible] present data suggest that a phonologic[illegible]aluation does not necessarily precede t[illegible] memory scan, and in fact, the two ope[illegible]ons may operate in parallel. It is hypo[illegible]zed that the RTs to the CCC + s are del[illegible] because there is a conflict of information[illegible] phonological evaluation leading to a n[illegible]ord response is opposed to the finding [illegible]eaning which might normally lead to a wo[illegible]esponse.

It is important to note that for [illegible] parallel model, no substantial changes need [illegible]e made regarding the nature of the operat[illegible]s themselves. It has usually been assume[illegible]hat the phonological evaluation involves s[illegible]ne rule process which analyzes and evalua[illegible] phonological lawfulness (Stanners et a[illegible], 1971). Rubenstein et al. (1971) suggest that part of this process also involves some ph[illegible]etic recoding of the visual stimulus. The [illegible]rch for meaning, on the other hand, has been described as a serial self-terminating scan through a subset of memory representations in search of some form of semantic marker (Stanners et al., 1971).

The postulation of a self-terminating scan for meaning was supported in several respects by the present data. First, the finding that RTs to CVCs were consistently longer than those to WORDs suggests that when the item was not a real word, *S* had to continue his search until he exhausted the set of representations he was searching. When meaning was found, however, the search was terminated, yielding faster RTs to WORDs. Secondly, a self-terminating scan leads to two related predictions concerning the RTs of incorrect responses. An error with a CVC would mean that a semantic marker was incorrectly detected, terminating the search. The RTs for CVC errors, therefore, should be shorter than those for correct CVC responses. This rela-

tionship should be opposite with WORDs, since an error indicated that *S* failed to find the marker, making the scan exhaustive. The data support these predictions: the CVC error RTs were 74 msec. less than the CVC correct responses, while the WORD errors were 79 msec. longer than the WORD correct responses. Similar results and conclusions were also reported by Stanners et al. (1971).

The parallel model not only has the advantage of being able to account for the CCC+ vs. CCC difference, but it is also not inconsistent with other aspects of the present data and previous data used to support the sequential model. As already mentioned, the finding that CVC RTs were longer than those to the other kinds of stimuli is easily predicted by the parallel model. Similar results were reported by Stanners et al. (1971), Rubenstein et al. (1970), and Snodgrass and Jarvella (1972), and this is probably the most consistent finding in this type of experiment.

The sequential model, as originally stated, clearly predicts that RTs to WORDs should be longer than those to CCCs, since with the latter stimuli, no memory search is required. This prediction was supported by Stanners et al. who found that RTs to CCCs were more than 115 msec. faster than those to WORDs. In the present study, however, RTs to WORDs essentially equalled those to CCCs. While this failure to replicate the previous findings may have been due to the fact that WORDs were responded to with the right hand, while nonwords were responded to with the left hand, it should be pointed out that at least two other studies have also reported that WORD RTs were as fast or faster than RTs to unlawful nonwords. Rubenstein et al. (1971) obtained RTs of 842 and 876 msec. for their words and unlawful nonwords, respectively, while Snodgrass and Jarvella (1972) reported RTs of 578 and 621 msec. for these types of stimuli, respectively. In both studies, the authors explain their data by asserting that when an item is phonologically unlawful, [illegible]preprocessing and the detection of that un-[illegible]lness, leading to a nonword response, [illegible]nger than the processing of lawful [illegible]

[illegible]ilarly, the parallel model can account for RTs of WORDs being slower or faster than those to CCCs by asserting that one of the operations takes longer than the other. It is also possible that the two operations may be differentially sensitive to such experimental manipulations as instructions or reward contingencies emphasizing either speed or accuracy. Therefore, it is clear that the relationship between the RTs to WORDs and CCCs is not critical for distinguishing between the parallel and sequential models, since both models can account for both kinds of data.

Experiment II

Experiment II was designed to control for the possibility that the memory search was sensitive not to meaning, but simply to the frequency of occurrence in the English language. Frequent CCCs were compared to infrequent CCCs much in the same way that CCC + s were compared to CCCs in Experiment I. If the memory search were sensitive to frequency, then, according to the present argument, RTs to the frequent CCCs should have been longer than those to the infrequent CCCs.

Method

Subjects. Twenty female undergraduate students enrolled in introductory psychology at the University of Connecticut served as *Ss* in the experiment. All of the *Ss* were experimentally naive. Each *S* received 1 hr. of experimental credit toward the course requirement.

Apparatus. The apparatus was the same as used in Experiment I.

Stimuli. The stimuli in the practice sessions were the same as in Experiment I. For the test trials, the 15 most frequent CCCs were selected on the basis of the total score in the norms reported by Underwood and Schulz (1960). The range of the frequency scores was 50 to 534, and the mean score was 153. From each of the frequent CCCs (CCCfs) another CCC was constructed by replacing the middle consonant with another consonant so that the resulting CCC had a frequency of 0 in the Underwood and Schulz norms. In addition, CVCs were constructed from each CCCf by inserting a random vowel in the center position (with the restriction that a word did not result). There were also 45 WORDs used in the test session, which were the same WORDs as used in the test session of Experiment I. Thus, there were again 90 test trials: 15 CCCfs, 15 CCCs, 15 CVCs, and 45 WORDS.

Procedure. All procedures were identical to those of Experiment I.

Results

All scoring and analyzing procedures were the same as in Experiment I. The results for the test trials appear in Table 2.

TABLE 2

Mean Reaction Times (RTs), Standard Deviations, and Percent Errors for the Test Trials in Experiment II

Stimulus type	Number of stimuli	Mean RT	*SD*	% error
CCCf	15	678	47	2.0
CCC	15	677	46	.7
WORD	45	685	49	1.4
CVC	15	792	75	7.6

Note. Abbreviations: C = consonant, V = vowel, f = frequent.

The analysis of variance revealed a significant treatment effect for the three nonword conditions, F (2, 38) = 49.11, $p <$ .001. A series of Duncan multiple range tests revealed that RTs to the CVCs were again greater than those to the other stimuli ($p < .01$), but there was no significant difference between CCCfs and CCCs, ruling out the frequency hypothesis. The lack of a CCCf vs. CCC difference in the overall means was paralleled by individual trends, and intrastimulus pair differences.

All other aspects of the results, including the RTs to WORDs, closely replicated the results of Experiment I.

Discussion

The hypothesis that *S*s decide that an item is a word or a nonword on the basis of a simple frequency determination was not supported by the present data. Although it might be argued that the CCCfs in the present study were not frequent enough to cause a delay in RT, it should be pointed out that they are more frequent than the CCC + s used in Experiment I, which did cause such a delay. Therefore, it seems likely that it was the meaning of the CCC + s, rather than their frequency, which caused the effect. As Rubenstein et al. (1970) suggest, the role of frequency might be to limit the subset of memory representations to be searched, but frequency itself is probably not the deciding factor.

Conclusion

The sequential model and the parallel model described above are based on many of the same arguments and assumptions, and can account for much of the same data. In Experiment I, however, the CCC vs. CCC+ comparison separates the predictions of the t[illegible]dels, and the data clearly support the [illegible] model. The results of Experiment II su[illegible]hat the memory search was not a sim[illegible]requency determination. The parallel mo[illegible]imilar in principle to other previous mod[illegible]ch as the one described by Ingalls (1972). [illegible] study, as in Experiment I, *S*s could ha[illegible]sponded on the basis of the prelimina[illegible]rations alone, if the organization of the o[illegible]ns was sequential. His data, however, s[illegible]ed that the higher order operations were [illegible]tilized to complete the task, without re[illegible] to the information gained by the prelimi[illegible]operations. Similarly, the most reasona[illegible]onclusion to be drawn from the present [illegible]s that *S*s are able to analyze the meaning o[illegible]timulus without first evaluating its p[illegible]logical properties.

References

Eichelman, W. H. Familiarity effects [illegible] simultaneous matching task. *Journal of [illegible]mental Psychology*, 1970, 86, 275–282.

Ingalls, R. P. The effects of sa[illegible]fferent patterns on tachistoscopic recognitio[illegible] letters. Unpublished doctoral dissertation, U[illegible]ity of Connecticut, 1972.

Krueger, L. E. Search time in a redund[illegible] visual display. *Journal of Experimental [illegible]hology*, 1970, 83, 391–399.

Norman, D. A. *Memory and attention.* [illegible] York: Wiley, 1969.

Novik, N., & Katz, L. High speed vis[illegible]anning of words and nonwords. *Journal of E[illegible]mental Psychology*, 1971, 91, 350–353.

Reicher, G. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, 81, 275–280.

Rubenstein, H., Garfield, L., & Millikan, J. A. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 487–494.

Rubenstein, H., Lewis, S., & Rubenstein, M. Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 645–657.

Snodgrass, J., & Jarvella, R. Some linguistic determinants of word classification. *Psychonomic Science*, 1972, 27, 220–221.

Stanners, R. F., Forbach, G. B., & Headley, D. B. Decision and search processes in word–nonword classification. *Journal of Experimental Psychology*, 1971, 90, 45–50.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

Wheeler, D. Processes in word recognition. *Cognitive Psychology*, 1970, 1, 59–85.

(Received September 20, 1973)

Journal of E[illegible]ntal Psychology
1974, Vol. [illegible] No. 6, 1021-1026

THE [illegible]MPORAL COURSE OF RECOVERY FROM INTERFERENCE AND DEGREE OF LEARNING IN THE BROWN–PETERSON PARADIGM [1]

DELOS D. WICKENS [2] AND M. M. GITTIS

The Ohio State University

In the Brown–Peterson paradigm, performance declines across successive trials with the same class of materials, but it improves as a function of the [illegible] value of an interposed intertrial interval. One interpretation of [illegible]fect is that the previous interfering items are forgotten with time. This interpretation was tested by repeating items prior to the increased intertrial [illegible] order to reduce their rate of forgetting. No evidence was obtained to indicate that this variable influenced the temporal course of recovery from [illegible]e interference, suggesting that this interpretation of the effect is inadequate.

[illegible] the Brown–Peterson paradigm (B[illegible]n, 1958; Peterson & Peterson, 1959), suc[illegible]ive trials with the same class of mat[illegible]ls are given at an intertrial interval in t[illegible] neighborhood of 30 sec., and with a con[illegible]nt retention interval for all trials, per[illegible]mance drops markedly and reaches asy[illegible]ptote in three or four trials (Keppel & [illegible]nderwood, 1962; Loess, 1967; Wickens, 19[illegible]). Two kinds of interpretations have be[illegible]n made of these data. One is that *S* be[illegible]omes bored or "inattentive" and hence performs less well. This interpretation seems most implausible, for the asymptote is reached within 2 min. after the experiment is begun, and if *S*'s attention were as fickle as this interpretation implies, one would have to regard as suspect the results of almost all laboratory research in psychology.

The most plausible interpretation of the decline is that lower performance on the later trials is engendered by the interference resulting from the prior inputs. Since an asymptote is reached, it must be assumed that the growth of interference is limited to some maximum achieved within about four trials. Obviously, this is a proactive interference interpretation. Such an interpretation does not state whether the deficit occurs in the encoding, storage, or retrieval stage, but simply that prior experience with the same class of items inhibits performance on subsequent items. The occasional occurrence of intrusions of prior items during the recall of later items offers some positive evidence of competition at the time of retrieval, although it does not negate an interference effect at the other stages.

The degree of decline in performance is a function of the intertrial interval (Cermak, 1969; Loess & Waugh, 1967); but what is more important for the present research is that if after n trials, an increase in intertrial interval occurs before trial $n+1$, performance will improve on trial $n+1$ (Peterson & Gentile, 1965).

Kincaid and Wickens (1970) found an orderly increment in trial $n+1$ as the critical intertrial interval was increased beyond the usual 30 sec. by 15–120 se[illegible] At this latter temporal value, performance was increased by about two thirds of the drop from Trial 1 to Trial 4. This general result was also obtained by Hopkins, Edwards, and Cook (1972).

A further interesting characteristic of this spacing effect is found in the research of Nield (1968). He used consonant trigrams (CCCs) as his target items and color naming as the rehearsal preventative task. During the lengthened intertrial interval, whose increment was 16, 48, or 80 sec. for different groups, his three sets of groups engaged in different types of

[1] The research was supported in part by National Science Foundation Grant GB-33680 to the first author.

[2] Requests for reprints should be sent to Delos D. Wickens, Department of Psychology, The Ohio State University, Columbus, Ohio 43210.

TABLE 1

Schematic Description of the Procedure for the Various Groups of the Experiment

Group	Trial sequence
Experimental 0	$C_0N_0C_1C_2C_3C_1C_2C_3C_4$
Control 0	$C_0N_0C_5C_6C_7C_1C_2C_3C_4$
Experimental 1	$C_0N_0C_1C_2C_3C_1C_2C_3N_1C_4$
Control 1	$C_0N_0C_5C_6C_7C_1C_2C_3N_1C_4$
Experimental 2	$C_0N_0C_1C_2C_3C_1C_2C_3N_1N_2C_4$
Control 2	$C_0N_0C_5C_6C_7C_1C_2C_3N_1N_2C_4$
Experimental 4	$C_0N_0C_1C_2C_3C_1C_2C_3N_1N_2N_3N_4C_4$
Control 4	$C_0N_0C_5C_6C_7C_1C_2C_3N_1N_2N_3N_4C_4$
Experimental 5	$C_0N_0C_1C_2C_3C_1C_2C_3N_1N_2N_3N_4N_5C_4$
Control 5	$C_0N_0C_5C_6C_7C_1C_2C_3N_1N_2N_3N_4N_5C_4$

Note. Subscripts indicate particular items (0 = practice item). Abbreviation: C = consonant trigram, N = number trigram.

activities. One set of groups simply rested, another set named a matrix of color patches, while a third set continued with the Brown–Peterson memory task, using three-digit numbers as the target items. This latter task was chosen because there seems to be little or no interference between CCC items and number trigram (NNN) items. Note also that the distractor task was color naming, not digit subtraction as in the original Peterson and Peterson (1959) experiment. The set of groups did not differ from each other either in the rate or magnitude of recovery from the inhibitory effect. Within the limits of the kinds of tasks employed—resting, color matrix reading, or the Brown–Peterson task with items of a different conceptual class, the passage of time alone seems to have been responsible for the recovery.

The phrase "within the limits of the kinds of tasks employed" should be emphasized, since there is evidence that the recovery effect is influenced by the nature of the interpolated activity between trial n and $n+1$ and the nature of the target activity. Evidence on this topic will be presented in the Discussion section.

Several possible ways of accounting for these results seem available. One is that they are due to a forgetting of the interfering material either because of mutual interference among these items—an "acid bath" (Posner & Konick, 1966) type of approach or to some time-related weakening factor or factors. Another interpretation is a list differentiation process made possible through the separation of the n and the $n+1$ trials by different intervening materials (Gorfein & Riley, 1974).

The present experiment was designed to test the forgetting-of-interference interpretation. It was assumed that if one could delay the forgetting of the interfering items, one would also depress the rate of recovery of performance across time. The method used to reduce forgetting of the interfering items was simply to give a dual presentation of the to-be-interfering items for one group and only a single presentation of the items for the other (control) group. This method is based simply on the time-honored belief that repetition increases the degree of learning and hence retards forgetting.

Method

Subjects. The *Ss* were 960 male undergraduate students at The Ohio State University who chose to participate in this experiment to fulfill a requirement for their introductory psychology course. The *Ss* were assigned to the 10 different groups in rotating order of their appearance at the laboratory, with 96 *Ss* in each group.

Material and apparatus. The target items of the experiment consisted of eight CCCs with association values between 25% and 33% (Witmer, 1935) and six three-digit NNNs. The NNNs were compiled from a table of random numbers such that no NNN had two identical digits or comprised a "meaningful" sequence (e.g., 321). The digit 7 was not used. Slides of the Stroop (1935) test were employed as the distractor or rehearsal preventative activity.[3] They consisted of 15 color names appearing in a 3 × 5 matrix. These were photographed from a sample given by Underwood (1966, p. 540). A plain 3 × 5 in. index card with a hand-printed sample of the Stroop test was shown to each *S* during the instructions. A metronome was used to produce clicks during the presentation of the Stroop slides. All slides were presented by a Kodak projector and timed by a 35-mm. timer.

Procedure. An asterisk presented for 2 sec. served as a ready signal for the 1.5-sec. presentation of the CCC or NNN which followed. Next was a 12-sec. presentation of a Stroop slide accompanied

[3] The Stroop test was chosen for the rehearsal preventative activity because we have, in The Ohio State University Laboratory, found it to be a very effective distractor task in the sense that *Ss* become quite involved and challenged by the task itself—more so than a number subtraction task. It seems to be highly effective for distracting *Ss* from rehearsing the memory task.

by 15 beats of the metronome, and finally there appeared a [illegible] presentation of a question mark. Each trial, therefore, took 21.5 sec., and each trial was exactly the same except for the target item presented and the specific Stroop slide given.

The design of the experiment contained two sets of groups, the experimental set and the control set. The experimental set experienced repetition; the control set did not. Table 1 presents the basic outline of the procedure. In that table the letter "C" stands for the consonant trigram, while the letter "N" represents a digit trigram. The subscripts designate particular items with the zero referring [illegible] practice items (C$_0$ and N$_0$) [illegible] experiment to accustom [illegible]

[illegible] Experimental Group 2, which is typical. It began with two practice trigrams followed by three trials each on a different CCC. For the next three trials these same trigrams were repeated in their original sequence. Then, with no break in tempo, Ss received two NNNs, and finally they were tested on the critical trigram C$_4$. Its control has a similar program except that no repetition is given. It will be noted that trigrams C$_1$, C$_2$, and C$_3$ occurred just before the shift to NNNs as they also did in the experimental group. The other groups differ in the number of NNNs inserted between C$_3$ and the critical item C$_4$. Since the intertrial interval was 21.5 sec., the time value between C$_3$ and C$_4$ can be determined by the number of interpolated N trials. The use of NNNs during the recovery period was based upon the Nield (1968) finding of no apparent difference between type of activity interpolated and the beneficial value of the spacing effect. In addition, dealing with subsequent memory load activities tended to prevent Ss from thinking about any prior items. And finally, the present experiment is not concerned with the precise form of the recovery function but with the comparison of different groups' recovery.

Results

The effect of repetition. Figure 1 shows the percent correct recall based on items correct per trigram for the experimental and control groups across the first six trials. An analysis of variance of these data showed the performance of the experimental group to be significantly superior to that of the control group on the repetition series of trigrams, F (1, 958) = 48.22, $p < .01$. The effect of the trial block (Trials 1–3 vs. Trials 4–6) was also highly significant, F (1, 958) = 3,002.12, $p < .01$, as was the interaction of the two main effects, F (1, 958) = 16,548, $p < .01$. These results are essentially a replication of the work of Cermak (1969).

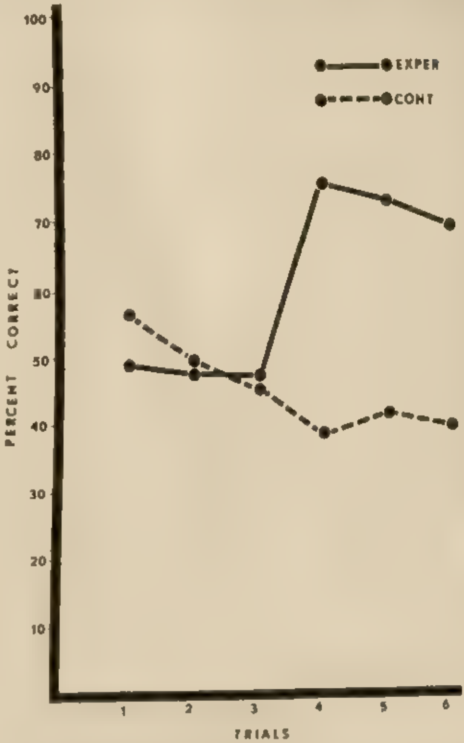


Figure 1. Performance of the experimental and control groups on the six items prior to the introduction of the increased intertrial interval.

It should be remarked that the usual decline in performance across the early trials does not occur at a significant level. Presumably this is because PI had already built up as a result of the practice trials, on which performance was not recorded. However there seems to be a slight downward trend across the first three or four trials for the control groups, showing the usual nadir of performance at about Trial 4.

The effect of temporal spacing on performance. The percent correct recall on the final trial as a function of the number of NNN slides interposed between the last presentation of a CCC and the critical trigram is shown in Figure 2. An analysis of variance of these data suggests that repetition had no effect, F (1, 950) = .285. There was no significant interaction between the effects of time and repetition, F (4, 950) = 1.928. The main effect of

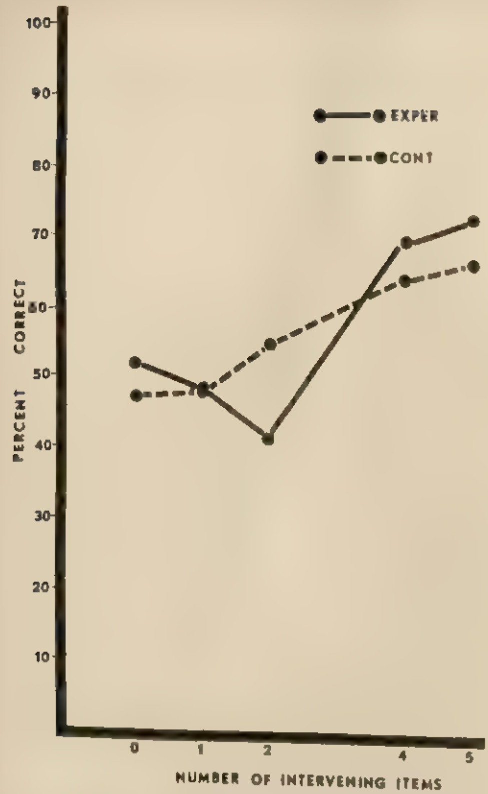


FIGURE 2. Performance on the new (critical) item as a function of the length of the added spacing interval. The duration of any intervening item presentation and test is 21.5 sec.

time, however, was highly significant, $F\ (4, 950) = 13.752$, $p < .01$.

Discussion

In beginning this discussion, it should be emphasized that the purpose of the experiment is not concerned with the retention of a previously presented target item, but rather with the measure of the effects of the strength of prior target items upon the recall of a different target item which is presented following a lengthening of the usual interval. Much data (Keppel & Underwood, 1962; Loess, 1967; Wickens, 1970) point to the fact that performance declines across the first three or four trials with the same class of materials. Presumably it does so because the input of the prior items interferes in some way with the performance on later items. With an increase in intertrial interval after trials with a constant interval, performance improves. The interference-from-prior [illegible] interpretation of the loss across tri[illegible]ld suggest that such gains would be du[illegible] forgetting of the prior items, and h[illegible]ecrement in interference and increment i[illegible]rmance.

This interpretation, there[illegible] would tend to account for the improvemen[illegible] performance which is associated with an [illegible]ed span of time between trial n and $n +$ [illegible] increased forgetting of the interferin[illegible]rial. The present experiment was desi[illegible] increase the retention of the interfe[illegible] items by repeating them in the earlier t[illegible]. Such an interpretation would predict [illegible]at better learned (and hence better reme[illegible]ered) items would persist longer in time [illegible] therefore would depress the temporal re[illegible]ery curve. The analysis of variance gave no [illegible]dence of a difference between the experime[illegible] and control groups, although the tim[illegible]fect was significant for both groups. De[illegible] the lack of main effect or interaction, the [illegible]o groups with two and five NNNs were c[illegible]pared by a t test, and they fell short of th[illegible] 05 level. Considering that there are 96 *Ss* in [illegible]ch group, the failure to find significance b[illegible]ween the experiments and controls is quite c[illegible]nvincing.

Assuming that the repetition did s[illegible]rengthen the first three triads, and Figure [illegible] indicates that it did, one should expect them t[illegible]e better retained across the interpolated ac[illegible]ity and, by producing greater interference, to suppress the recovery curve. Obviously this did not occur, and so no support for the forgetting of the interfering material is found

Research on the present topic by Hasher, Goggin, and Riley (1973) and also by Goggin and Riley (1974) led the authors to a list differentiation interpretation of the effect. In their research, they worked with word triads as their target items, and after presenting three triads all drawn from the same category, they (*a*) shifted to a new category for three trials; (*b*) continued with triads of the same category for three trials; or (*c*) shifted to a rest condition in which *Ss* viewed pictures. On the seventh trial they reverted to a triad from the original category. Enhancement of performance on the seventh trial was found only for the rest groups, despite the fact that the group shifted to a new category on Trial 4 showed release from PI (Wickens, 1970).

The interpretation they made of these results was that the interposed highly different activity of "rest" permitted or generated a differentiation between the specific materials of the first three trials and the material of the

seven[illegible] even though all four items came from [illegible] ame category. By so doing, the interf[illegible] of the previous three items was reduc[illegible] during seventh-item recall. Words in gene[illegible] even though of a different category, are [illegible] able to produce the necessary psycholog[illegible] contrast for differentiation.

[illegible] position that the differentiation was due [illegible] ply to the change in the nature of the task [illegible] n be contradicted for several empirical reasons. It will be noted in the Nield (1968) experiment described above that the group that continued to perform the Brown–Peterson task, but on a new class of materials—NNNs [illegible] ther than CCCs—demonstrated no differ[illegible] from the groups which either color named [illegible]

[illegible] cond experiment, Nield (1968) ch[illegible] the task and did or did not change the na[illegible] of the materials. Half of his groups be[illegible] their Brown–Peterson trials with CCCs as [illegible] target items and half with NNNs. Aft[illegible] three trials, they were shifted to a new task which consisted of simply reading a ma[illegible]x of numbers or of consonants. It had bee[illegible] clearly impressed upon *S*s that this shift wo[illegible]d occur and that they were not to be tested on the recall of the matrix material. Obviously, for half of each group the matrix was the same material as the target items, and for half it was different. On the subsequent Brown–Peterson trial, *S*s were tested with the class of materials of the original three triads. If the materials were the same type throughout the Brown–Peterson, matrix, Brown–Peterson sequence, no enhancement of performance on the final trial was formed. If the Brown–Peterson material and matrix material differed, however, performance on the final trial was markedly facilitated. Clearly, simply changing the task from a memorial one to a simple processing one does not achieve the psychological contrast effect which is necessary for list differentiation. It is apparent that, in our present state of knowledge, it is difficult to predict what manipulation will produce differentiation (if that is the basis for the improvement); although we seem to know that the employment of the same general class of material throughout, even though in a different manner, will prevent it.

An example of the difficulty in predicting when differentiation will make itself felt is found in the results of the present experiment. Underwood (1969) has presented a considerable body of evidence to the effect that frequency is a significant attribute of memory, and *S*s are highly capable of responding to differences in frequency of presentation independent of response strength itself. Our experimental group received two presentations of the three pre-interpolated-task triads, and the post-interpolated triad received only one. The ratio of 2:1 should have established a differentiation between the first set of triads and the test triad. This frequency differential would not have been available to the control group. Hence, one should predict that the experimental group would excel the control group, but it did not do so.

In conclusion it would appear that the forgetting-of-interfering-materials concept has no support from the data of this experiment; and that a differentiation interpretation, even though it is not without problems, may be somewhat more promising.

REFERENCES

Brown, J. A. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

Cermak, L. S. Repetition and encoding in short-term memory. *Journal of Experimental Psychology*, 1969, **82**, 321–326.

Goggin, J., & Riley, D. A. Maintenance of interference in short-term memory. *Journal of Experimental Psychology*, 1974, **102**, 1027–1034.

Hasher, L., Goggin, J., & Riley, P. A. Learning and interference effects in short-term memory. *Journal of Experimental Psychology*, 1973, **101**, 1–9.

Hopkins, R. H., Edwards, R. E., & Cook, C. L. The dissipation and release of proactive interference in a short-term memory task. *Psychonomic Science*, 1972, **27**, 65–67.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Kincaid, J. P., & Wickens, D. D. Temporal gradient of release from proactive inhibition. *Journal of Experimental Psychology*, 1970, **86**, 313–316.

Loess, H. Short-term memory, word class, and sequence of items. *Journal of Experimental Psychology*, 1967, **74**, 556–561.

Loess, H., & Waugh, N. C. Short-term memory and intertrial interval. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 455–460.

Nield, A. F. The effects of time and activity on the dissipation of proactive inhibition in short-term memory. Unpublished master's thesis, The Ohio State University, 1968.

Peterson, L. R., & Gentile, A. Proactive interference as a function of the time between tests. *Journal of Experimental Psychology*, 1965, **70**, 473–478.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

Posner, M. I., & Konick, A. F. On the role of interference in short-term retention. *Journal of Experimental Psychology*, 1966, **72**, 221–231.

Stroop, J. R. Studies in interference in serial verbal reactions. *Journal of Experimental Psychology*. 1935, **18**, 643–662.

Underwood, B. J. *Experin psychology*. New York: Appleton-Century-(1966.

Underwood, B. J. Attributes of y. *Psychological Review*, 1969, **76**, 559–57

Wickens, D. D. Encoding cat of words: An empirical approach to m *Psychological Review*, 1970, **77**, 1–15

Witmer, L. R. The association val hree-place consonant syllables. *Journal netic Psychology*, 1929, **4**, 31–55.

(Received September 21

Journal of Exp[illegible] Psychology
1974, Vol. 10[illegible] 1027–1034

MAINTENANCE OF INTERFERENCE IN SHORT-TERM MEMORY[1]

JUDITH GOGGIN[2]
University of Texas at El Paso

DONALD A. RILEY
University of California, Berkeley

The nature of facilitation and interference effects was examined within the Brown–Peterson paradigm. Subjects were first given a block of three tests on items from a single taxonomic category. Three variables dictated the sequence of events thereafter: (a) the similarity relationship between a subsequent block of items and Block 1 items, i.e., same category–same items, same category–different items, or different category; (b) the time interval between Block 1 [illegible]uent block of interest, i.e., immediate vs. delayed presentation; [illegible]hin the delayed presentation conditions, whether the interval was [illegible] memory task or with rest. Recall was best when items were [illegible] with different items from the same category, and intermediate [illegible] the category changed. A rest delay improved recall in all conditions [illegible] both to no delay and to a delay filled with work. A differentiation [illegible]othesis was used to account for these results.

Recently Hasher, Goggin, and Riley (1973) demonstrated that both facilitative and inhibitory effects occurring within the Brown–Peterson paradigm (Brown, 1958; Peterson & Peterson, 1959) persist over intervals that would normally be considered outside the range of short-term memory (STM). In that experiment, *S*s were first tested with three triads of words from one taxonomic category, then three triads from another category, and finally three more triads from the first taxonomic category. The time intervening between the first and the third sets was approximately 90 sec. When the third set of triads was identical to the first set, facilitation over an appropriate control was demonstrated, thus showing a memorial effect for specific items. When, on the other hand, the third set of triads was from the same category as the first but was comprised of different words, inhibition equal to a control tested on the same category throughout was demonstrated. Controls made it clear that the inhibition was category specific.

Previous research (e.g., Cermak, 1969; Hebb, 1961) has shown the facilitative effects of repetition over rather long intervals in an STM paradigm, but studies investigating the persistence of inhibition have shown considerable loss of inhibitory effects. That is, for comparable time intervals, facilitation persists, but inhibition is lost. Some studies (e.g., Hopkins, Edwards, & Cook, 1972; Kincaid & Wickens, 1970) have found a moderate amount of recovery (approximately 62% and 48%, respectively) from the build-up of proactive interference (PI) after a 90-sec. interval, and Cermak (1970) found almost complete recovery following a 66-sec. rest interval. In the Hasher et al. (1973) study, on the other hand, such recovery was not found.

We speculated that the critical difference between our study and previous ones lay in the nature of the interpolated material and its effects on *S*'s ability to differentiate clearly between original learning and later learning. Only in Hasher et al. (1973) were the interpolated material and task identical in nature to those of original and final learning. The *S*s were required to recall items different only in that they came from a different taxonomic category. The other experiments used different kinds of activities in the intertest interval, such as symbol cancellation, Stroop color naming, and rest, but in no case were *S*s engaged in a memory task of the same nature. To determine whether or not the differences between these two classes of experiments are related to the nature of the intervening activity, it is necessary to include within the same design conditions that have a memory task as interpolated activity and conditions that

[1] This study was supported by National Science Foundation Grant GB-36210 to the first author.

[2] Requests for reprints should be sent to Judith Goggin, Department of Psychology, University of Texas at El Paso, El Paso, Texas 79968.

TABLE 1

Experimental Conditions

Condition	Block 1 (Tests 1–3)	Block 2 (Tests 4–6)	Block 3 (Tests 7–9)
1	C_1	rest	C_{1s}
2	C_1	C_2	C_{1s}
3	C_1	C_{1s}	rest
4	C_1	rest	C_{1d}
5	C_1	C_2	C_{1d}
6	C_1	C_{1d}	C_{1d}
7	C_1	rest	C_3
8	C_1	C_2	C_3

Note. Abbreviations: C = taxonomic category, s = repetition of the same items, d = use of different items from the same category.

have some other sort of nonmemorial filler task. This was the purpose of the present experiment.

Method

Design. Outlined in Table 1 are the eight experimental conditions, each one of which was divided into three temporally equivalent blocks of learning or rest activity. A block of learning activity consisted of three Brown–Peterson tests; all to-be-remembered items in any one block belonged to the same taxonomic category (C). Numerical subscripts represent category membership. Thus, Condition 8 received a different category in each of the three blocks (C_1, C_2, C_3), while all items belonged to the same category in Condition 6 (C_1, C_1, C_1). When there was category repetition across blocks, letter subscripts designate the degree of item similarity. Repetition of the same items is represented by the letter "s"; the use of different items from the same category is represented by the letter "d." This distinction can be illustrated by an examination of Conditions 1 and 4. If the words on the first test of Block 1 had been, for example, MUSKRAT, WOLF, ELK, these same words would again appear on the first test of Block 3 in Condition 1, while GOAT, MULE, CAMEL might appear in the corresponding position in Condition 4.

All conditions were treated alike in the first block, receiving three STM tests on items from a single category. Dictating the sequence of events thereafter were three main variables: (*a*) the similarity relationship between a subsequent block of items and Block 1 items, i.e., same category–same items (Same), same category–different items (Different), or different category (Release); (*b*) the time interval between Block 1 and a subsequent block of interest, i.e., immediate or delayed presentation; and (*c*) within the delayed presentation conditions, whether the interval was a work interval filled with a memory task or a rest interval in which *S*s looked at scenic slides. Consider, for example, the same category–same item relation. Both [illegible] ons 1 and 3 received C_{1s}, but C_{1s} was t[illegible] ter a delay in Condition 1, whereas it was [illegible] mmediately in Condition 3. Comparison o[illegible] mance on C_{1s} can also be made between [illegible] ns 1 and 2, which were identical except th[illegible] lay was a rest interval in the former and wa[illegible] ith tests on a new category in the latter. Pa[illegible] mparisons on same category–different items [illegible] ible in Conditions 4, 5, and 6. C_{1d} was tes[illegible] r no delay in Condition 6, after a rest dela[illegible] dition 4, and after a work delay in Condi[illegible] Finally, performance on a new category c[illegible] ssessed after no delay (Conditions 2 and 5 [illegible] rest delay (Condition 7), and after a wor[illegible] (Condition 8).

Materials. Foods, body parts, and animals were the three taxonomic categories from which stimuli were chosen. The categories were used equally often within each condition and occurred equally often in each of the three blocks of trials. Each category was composed of 36 words, divided arbitrarily into 12 triads. Assignment of the 36 triads to conditions and test positions within conditions could not be equalized, but was reasonably well balanced. The 3 words in each triad were arranged diagonally across a slide.

Procedure. Subjects were first given practice trials to familiarize them with the task. Practice trials were identical to regular test trials except that symbols, instead of words, were the to-be-remembered items. Each *S* then received three 90-sec. blocks of learning or rest activity. A block of learning activity consisted of three Brown–Peterson tests, each of which lasted 30 sec. The sequence of events during this 30-sec. period was as follows: (*a*) a ready signal for 2.5 sec.; (*b*) a word triad presented for 1.5 sec., during which time *S* read the words aloud; (*c*) a 15-sec. presentation of a three-digit number from which *S* was instructed to count backward by threes as fast as possible; and (*d*) a question mark displayed for 11 sec., which was *S*'s signal to recall the word triad, in order if possible. If another test was to follow, it began immediately.

During blocks of rest, *S*s were shown 12 scenic slides. These slides were presented for the same time intervals (i.e., 2.5, 1.5, 15, and 11 sec.) that would have been used had the block been filled with STM tests. The *S*s were instructed simply to relax and to view the pictures. If a block of learning activity followed a block of rest, *S*s were forewarned shortly prior to the appearance of the ready signal. Slides were presented by a Carousel projector, with slide duration regulated by a Gerbrands tape timer.

Subjects. A total of 384 *S*s was used in the experiment, with 48 in each condition. The *S*s were students at the University of Texas at El Paso who participated in the study to receive bonus points in introductory psychology. The *S*s were randomly assigned to conditions in blocks of 8. Thirteen *S*s were discarded for failure to follow instructions or equipment malfunction.

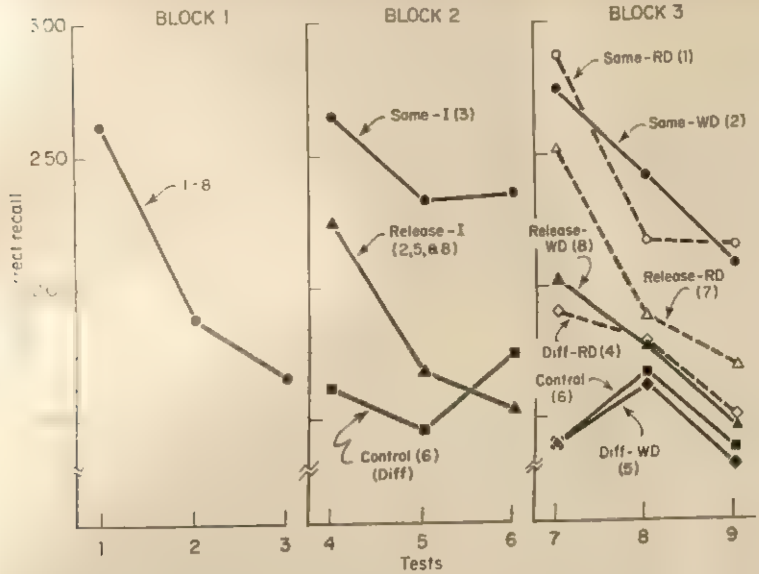


FIGURE 1. Mean correct recall as a function of condition and test. (Abbreviations: I = immediate presentation, RD = rest delay, WD = work delay, Diff = different.)

RESULTS

An item was considered correct if it was given, whether or not it was placed in correct serial position. Mean correct recall is presented in Figure 1. The figure is divided into three panels that correspond to the three blocks of trials. All conditions were treated alike in Block 1 and, consequently, are shown as a single curve. Those conditions treated alike in both Blocks 1 and 2 were similarly collapsed in Block 2 for simplicity of exposition. None of these conditions so treated differed significantly from each other.

Replication of basic phenomena. High recall on Test 1, followed by a drop in retention with succeeding tests on similar stimuli, is characteristic of the Brown–Peterson paradigm. The rapid drop in recall, which has been attributed to increasing PI across tests, is evident in Panel 1. This main effect of tests was significant, F (2, 752) = 140.93, $p < .001$; no other effects were.

Effects of block similarity. In the second and third blocks of work activity, there were three types of items defined by their similarity relationship to the Block 1 items: Items were drawn from the same category (Different–Control), were drawn from a new category (Release), or were repetitions of items that had appeared in Block 1 (Same). The effect of the similarity variable was examined on Test 4, across Block 2, and across Block 3.

On Test 4, the variations in item similarity had a strong and reliable effect on the level of recall, F (2, 141) = 16.52, $p < .01$. After PI has built up within a category, a change to a new category produces release from interference as evidenced by a marked improvement in recall (cf. Wickens, 1970). Replication of this phenomenon can be seen by comparing Release and Control conditions; a planned comparison showed that this difference was significant, F (1, 141) = 12.18, $p < .01$. Test 4 analyses also showed that repetition facilitated recall since Same *Ss* performed better than Release *Ss*, F (1, 141) = 4.88, $p < .05$.

Performance across the three tests of Block 2 is displayed in Panel 2. The general trend of recall by the Release condition was similar to that found in Block 1. On the first test with the new category, recall was high; on succeeding tests with similar

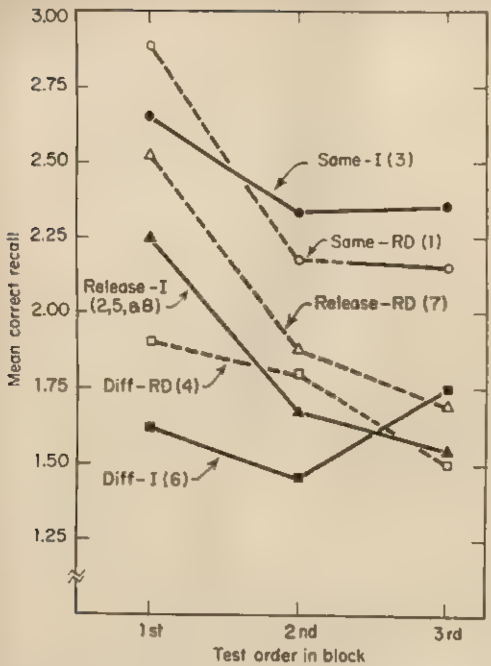


FIGURE 2. Mean correct recall as a function of condition and intertest interval. (Abbreviations: I = immediate presentation, RD = rest delay, Diff = different.)

materials, recall decreased, presumably because of cumulating interference. Thus, Control and Release *S*s, whose recall was quite discrepant on Test 4, were performing at similar levels on Tests 5 and 6. Performance in the Same condition showed the benefit of item repetition, with recall being maintained at a high level throughout the three tests. Analysis of Block 2 showed that both conditions, F (2, 141) = 25.75, $p < .001$, and tests, F (2, 282) = 6.60, $p < .01$, produced reliable differences. In addition, the interaction between conditions and tests was significant, F (4, 282) = 2.78, $p < .05$.

The effects of item similarity are somewhat more difficult to see in Block 3 because there were two groups for each similarity condition. The two groups are distinguished on the basis of whether Block 2 consisted of a rest interval or a work interval. Nevertheless, the general picture presented in Panel 3 is similar to that in Panel 2. The differences among similarity conditions occurred again and were significant, F (2, 282) = 54.96, $p < .001$. Overall, Same *S*s showed the benefit of item repetition; Different *S*s showed the depressing effect of cumulative PI; and Release *S*s showed the facilitation that accompanies a category change. Recall decreased over tests in the block, F (2, 564) = 25.88, $p < .001$. This main effect, however, must be interpreted in conjunction with the reliable Conditions × Tests interaction, F (4, 564) = 3.[illegible], $p < .05$. Recall decreased more rapidly over tests in the Same and Release conditions than in the Different condition, apparently reflecting some lower limit on recall in the last condition.

In summary, variations in item similarity between blocks of tests produced strong and consistent differences in recall. These differences were similar to those found by Hasher et al. (1973). After PI had built up within a taxonomic category, recall was depressed if there was continued testing on that category, or even if there was later return to that category after an interval of time. If there was a change in encoding category (Release), on the other hand, recall improved on the first test with the new items and then decreased as more items were presented in that category. Finally, repetition of items (Same) produced an improvement in recall in comparison to the release conditions, indicating that specific items were retained in this paradigm.

Effects of intertest interval. It has been pointed out that triads used in Blocks 2 and 3 were related to Block 1 triads in three ways, i.e., items were repeated, came from a new category, or were different items from the same category. One question of interest was whether performance on these later tests would be affected by the time interval separating them from Block 1 tests. This can be determined by comparing the level of recall for Same, Release, and Different conditions tested immediately with that after rest delay. Figure 2 shows this comparison.

Recall was better after a rest interval than after no delay on the first test of the block, F (1, 282) = 7.42, $p < .01$. Although the individual similarity conditions

TABLE 2

FREQUENCY AND SOURCES OF INTRUSIONS

…sions	Condition							
	1	2	3	4	5	6	7	8
Err… block	38.0	38.0	45.0	43.5	43.7	51.3	39.5	31.7
P…tion from each source								
…alist	.95	.88	.92	.85	.89	.88	.71	.84
…me category	1.00	.96	1.00	1.00	.93	1.00	1.00	.88
Test …	.79	.77	.72	.61	.68	.63	.80	.78
[illegible]	.14	.16	.28	.14	.15	.17	.20	.19
[illegible]	.07	.06	—	.26	.17	.20	0.0	.04

did [illegible] …tistically reliable effect, the [illegible] of the effect in each of the three [illegible] plus the absence of any relia… [illegible] (F = .02) makes it clear that a [illegible] facilitated recall in each condition. …erior recall after a delay has been [illegible] by others under similar conditions [illegible]. Different *S*s (Cermak, 1970; Kinca… & Wickens, 1970) and, in one case, for bo… Different and Release *S*s (Hopkins et al., 1972).

Delay did not have a reliable effect when the block was considered in its entirety. The interaction between delay and tests was, however, reliable, $F(2, 564) = 3.06$, $p < .05$. As Figure 2 shows, the facilitation by delay was maintained across tests in the Release condition, but not in the other two.

Effects of interval filler. The main purpose of the present experiment was to determine whether the nature of the material interpolated between the original learning task (Block 1) and the final learning task (Block 3) would affect the level of recall in the third block of tests. These data are displayed in Panel 3 of Figure 1.

Condition 6 was not included in the overall analyses of interval filler, but it was compared to Condition 5 in a separate analysis. These groups did not differ statistically in Block 3 ($F < 1$), and their overall performance was strikingly similar. Thus, *S*s returning to an old category with new items after an intervening different category (Condition 5) performed the same as *S*s maintained throughout on the old category (Condition 6), i.e., inhibition did not dissipate with time even though release from this inhibition had occurred in Block 2. Similar results were found in Hasher et al. (1973).

Overall analyses of filler activity were conducted both on the first test of the block and on the block as a whole. On Test 7, recall was higher when Block 2 was filled with rest than when filled with work, $F(1, 282) = 14.68$, $p < .01$. Preplanned comparisons showed a significant effect of filler activity for both Different and Release *S*s, Fs $(1, 282) = 8.70$, ps $< .01$, but for Same *S*s, this effect, while in the same direction, was not significant ($F < 1$). The interaction between condition and filler activity was not reliable.

Examination of Panel 3 shows that these effects persisted throughout Block 3. Recall was higher when a rest interval occurred between the first and third blocks than when the interval was filled with tests on a different category, $F(1, 282) = 7.82$, $p < .01$.

Despite this pattern of results, the interaction between condition and filler did not reach significance, making conclusions about the differential effect of filler on similarity conditions ambiguous. Again, preplanned comparisons showed that recall after a rest interval was higher than after a work interval for Different and Release conditions, Fs $(1, 282) = 19.46$ and 18.52, ps $< .01$, respectively, but a difference in the Same condition was not found ($F < 1$).

Error analyses. A breakdown of the incorrect responses is shown in Table 2. The measure of errors/block was used because conditions differed in the number of blocks of learning activity. The most

intrusions were made in Condition 6, in which all items were different but belonged to the same category; the fewest intrusions occurred in Condition 8, in which a different category of items was learned in each block. This would be expected if number of intrusions reflected the amount of interference.

Most of the errors came from within the list; most of these, and in some cases all, were intrusions from an appropriate category. Thus, *S*s appeared to be able to monitor their recall and to restrict their output to the correct taxonomic category. By far the largest percentage of intralist errors, 72%, originated from the immediately preceding test trial. Frequency of intrusions from earlier test trials was not high, although in three groups some errors were Block 1 intrusions into Block 3 recall. These percentages were 14%, 15%, and 6% for Conditions 4, 5, and 6, respectively. It appears that even after 90 sec. of rest or other learning activity, Block 1 items are available to occur as intrusions if they are of the appropriate category. Frequency of this kind of error in the other conditions was almost nonexistent.

Discussion

The present study was addressed to three issues arising from differences in the results obtained by Hasher et al. (1973) and previous investigators. First, of the various studies that have examined the persistence of proactive interference in the Brown–Peterson paradigm over time, only Hasher et al. found no reduction in interference. That study, however, did not have an appropriate nonwork condition. Consequently, one cannot be sure that loss of inhibition would have been found under conditions comparable to other experiments. This study provides such a comparison. Second, the present study allows a direct comparison between a retention interval filled with rest as opposed to further rehearsal of words from a different category and further rehearsal of words from the same category. This comparison is necessary in order to make a direct connection between the presence or absence of persisting inhibition and the nature of the filler activity. Third, this experiment, by extending the question of the effects of the interpolated activity to release and repetition of original learning conditions, allows a test of the PI decay hypothesis.

Consider first the effects of time. In the Results section, these effects were analyzed both on the first trial and on all three trials of the relevant blocks. While the latter measure may be of interest, it is only the first-trial data that give an unambiguous measure of the effect of previous experimental operations. Consequently, our emphasis here will be on this measure. Figure 2 showed that our Different–Delayed condition was reliably superior to the Different–Immediate condition. Thus, we replicated the results of previous experimenters showing that PI dissipates with time. This facilitative effect of delay occurred, and to the same degree, in the Release and Same conditions.

The findings of Hasher et al. were reproduced in this study in that there was no such dissipation of PI in the Different condition if the interval was filled with tests on items from another category. Filling the interval also depressed recall in the Release condition; while this effect appeared in the Same condition, it was less reliable. It is of interest to observe that the facilitative effects on recall of time and those associated with a change in category are additive. That is, the superiority of recall following a rest interval as opposed to a work interval was as great in the Release condition as it was in the Different condition. The absence of an interaction between category change and intertest interval can also be seen in Hopkins et al. (1972).

It has been suggested (e.g., Kincaid & Wickens, 1970) that PI decreases as a function of intertest interval because there is a time-dependent decay of the encoding characteristics responsible for PI. Several aspects of the present data argue against such an interpretation. Presumably, a change in the type of items increases recall because there is a release from the interference that has accrued for the original encoding category (Wickens, 1970). If this is the case, a rest interval should be of no particular benefit when it is followed by a category change. That is, since the items following a change in category are not encoded in the same way as the original items, decay of the encoding characteristics of the original items associated with a long intertest interval should be irrelevant. A rest interval in the present study, however, facilitated recall just as much when there was a category change (Release) as when there was no category change (Different).

There is another feature of these data that weighs against a time-dependent decay interpretation of the effects of delay. One set of *S*s (Sample [illegible]) received a later repetition of the original block of items; in this case, recall also improved as a function of interblock interval. It seems somewhat unreasonable to suppose that retention of items on their second presentation was facilitated because there was more forgetting of the items from their first present[illegible]

[illegible] en established by this [illegible] of other investigators: (*a*) PI is reduced [illegible] ing a rest interval; and (*b*) PI is not reduced when the interval is filled with tests on a different taxonomic category. There is a third fact, however, that must be added in order to complete the picture. Several experiments (Cermak, 1969; Nield, 1968; Wickens & Gittis, 1974) have demonstrated that after PI has built up in a Brown–Peterson design with consonant trigrams as stimulus materials, filling the intertest interval with tests on number triads does *not* retard the reduction of PI. That is, with CCC stimuli, there is no apparent difference in performance on the next retention test whether the interval is unfilled or whether it is filled with tests on NNNs.

One explanation of these three facts must be ruled out. The task used as an interblock interval filler cannot be responsible for the dissipation or the maintenance of PI in the studies under consideration. Although Nield's (1968) results, for example, showed dissipation of PI over a filled interval and the present study did not, both procedures were ones in which *S* engaged in a Brown–Peterson learning task during the interval.

An alternative hypothesis, more consistent with the present data, is that the amount of PI depends on the differentiation that has occurred during the intertest interval. Two types of differentiation can be distinguished, the first being dependent on time, the second, on item similarity. An unfilled interval allows facilitation of recall, not because of a weakening of Block 1 items, but because the time gap allows them to be distinctively identified as coming from a different part of the experiment. The pool of items from which *S*s must retrieve the correct items during Block 3 will tend to be restricted to those encountered recently. Proactive inhibition is thus reduced because the Block 3 items have a new, unique encoding cue to distinguish them from Block 1 items. If, on the other hand, the interblock interval contains more of the same memory task with similar material, temporal differentiation cues are eliminated. Even though Block 3 is separated from Block 1 by the same nominal time whether the interval is filled with work or rest, the functional time is different.

Why should materials that are sufficiently different from Block 1 items to allow release from PI (i.e., items from a different taxonomic category) produce such an effect? Perhaps as *S* tries to recall Block 2 items, he must repeatedly examine and reject erroneous intrusions from Block 1. This activity may in turn bridge the temporal gap between Blocks 1 and 3. When, however, letters are the to-be-remembered items and numbers are learned during the interval (e.g., Wickens & Gittis, 1974), the items are sufficiently different in kind that letters do not intrude while *S* is trying to recall numbers. In these conditions, Block 1 items do not intrude and need not be rejected during Block 2, temporal differentiation between Blocks 1 and 3 occurs, and loss of PI in Block 3 is observed. The greater release of PI from Block 1 to Block 2 with letters and numbers (Wickens & Gittis, 1974) than with different taxonomic categories as in the present study is consistent with this interpretation.

REFERENCES

Brown, J. A. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

Cermak, L. S. Repetition and encoding in short-term memory. *Journal of Experimental Psychology*, 1969, **82**, 321–326.

Cermak, L. S. Decay of interference as a function of the intertrial interval in short-term memory. *Journal of Experimental Psychology*, 1970, **84**, 499–501.

Hasher, L., Goggin, J., & Riley, D. A. Learning and interference effects in short-term memory. *Journal of Experimental Psychology*, 1973, **101**, 1–9.

Hebb, D. O. Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning*. London and New York: Oxford University Press, 1961.

Hopkins, R. H., Edwards, R. E., & Cook, C. L. The dissipation and release of proactive interference in a short-term memory task. *Psychonomic Science*, 1972, **27**, 65–67.

Kincaid, J. P., & Wickens, D. D. Temporal gradient of release from proactive inhibition.

Journal of Experimental Psychology, 1970, **86**, 313–316.

NIELD, A. F. The effects of time and activity on the dissipation of proactive inhibition in short-term memory. Unpublished master's thesis, Ohio State University, 1968.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

WICKENS, D. D. Encoding cat[illegible]s of words: An empirical approach to meani[illegible] *Psychological Review*, 1970, **77**, 1–15.

WICKENS, D. D., & GITTIS, M. [illegible]e temporal course of recovery from interf[illegible] and degree of learning in the Brown–P[illegible] paradigm. *Journal of Experimental Psyc[illegible]* 1974, **102**, 1021–1026.

(Received December 3,

Journal of Exp[illegible] Psychology
1974, Vol. 10[illegible] 1035–1038

INT[illegible]TION OF POSITIVE AND NEGATIVE LABELS WITH CA[illegible]RY COMPOSITION IN ATTRIBUTE IDENTIFICATION CONCEPT PERFORMANCE [1]

THOMAS C. TOPPINO AND PEDER J. JOHNSON [2]

University of New Mexico

The effect of response label (example or nonexample) was separated from the effect of ca[illegible] composition (the stimuli contained in a conceptual category) [illegible]s of complementary concepts and requiring *S* to solve two [illegible] with only examples or only nonexamples. It was pre[illegible] positive label (example) would facilitate performance in cate-[illegible] conditions characterized by a high relative frequency of [illegible] stimulus values, while the negative label (nonexample) would facili-[illegible]rformance in category composition conditions characterized by a low [illegible]e frequency of relevant stimulus values. Results on the initial problem confirmed the experimental prediction. By the second problem, the effects of response label had disappeared, leaving only a category composition effect. The experimental hypotheses were discussed in relation to an earlier hypothesis based on preexperimental experience.

A long-standing generalization in the concept literature is that instances from the positive category (examples) lead to more rapid solution than instances from the negative category (nonexamples). This generalization gained considerable support from early attribute identification (AI) concept experiments (rule known, relevant stimulus values unknown) where *S* was required to solve a given concept by using instances from only the positive or only the negative category (Freibergs & Tulving, 1961; Hovland & Weiss, 1953). However, recent studies (Davidson, 1969; Toppino & Johnson, 1973) have shown that the earlier experiments confounded two potential sources of solution difficulty: *category composition* (the stimulus types composing the category, i.e., TT, TF, FT, and/or FF)[3], and *response label* (whether the category is called positive or negative). In short, the positive and negative categories for any given concept differ in two respects: They contain different members of the stimulus population, and they are assigned different valenced labels. Thus, differences in performance associated with positive and negative instances of a given concept could be due to either or both of these factors.

The independent effects of category composition and response label can be experimentally isolated by using a pair of complementary concepts for which the positive category of each concept has the same composition as the negative category of its complement (Davidson, 1969). When the effects of category composition and response label are disentangled in this manner, results indicate that a positive label does not necessarily facilitate performance, and that solution difficulty depends primarily upon category composition factors (Davidson, 1969; Toppino & Johnson, 1973). However, response label is not completely without effect. In each experiment, there was at least one condition in which positive superiority was obtained independently of category composition. These findings suggest a weak form of the early positive superiority generalization, viz., while response label does not always influence performance, when it does, solution is easier

---

[1] [illegible]is research was supported by National Institute [illegible] Child Health and Human Development [illegible] HD-03390-06 awarded to the second author. T[illegible]thors wish to thank Brian Euler for assisting w[illegible]e collection of the data.

[2] R[illegible]quests for reprints should be sent to Peder J. Johnson, Department of Psychology, University of New Mexico, Albuquerque, New Mexico 87131.

[3] Any stimulus population can be partitioned into four stimulus types (TT, TF, FT, and FF) depending upon the joint presence (T) and/or absence (F) of two relevant values (Haygood & Bourne, 1965).

with a positive label. The purpose of the present experiment was to investigate the possibility that, under certain category composition conditions, a negative response label may result in superior AI performance.

Toppino and Johnson (1973, Experiment I) employed two pairs of complementary concepts (inclusive disjunction/joint denial and biconditional/exclusive disjunction) resulting in a positive and a negative label condition within each of four category composition conditions [(TT, TF, FT), (FF), (TT, FF), and (TF, FT)]. The results indicated that a positive label yielded a clear advantage only when category composition had a high relative frequency of relevant values [i.e., (TT, TF, FT)].[4] Furthermore, the advantage of a positive label relative to a negative label, appeared to decrease as the relative frequency of relevant values decreased across category composition conditions [i.e., (TT, TF, FT) > (TF, FT) > (TT, FF)].

The above findings suggest that a positive label somehow facilitates or encourages the use of relative frequency cues. Since the label "example" seems to connote presence of something, a positive label may encourage *S* to use an *affirmational strategy* whereby stimulus values that are present in the training instances become the basis of *S*'s pool of hypotheses. Such a strategy would facilitate performance to the extent that the relevant values occur with a high relative frequency. However, a negative label may have the opposite function. Since the label "nonexample" seems to connote the absence of something, a negative label may encourage *S* to use a *negational strategy* whereby stimulus values that are present in the training instances are excluded from *S*'s pool of hypotheses. This would predict that a negative label should lead to superior performance in the (FF) category composition conditions where the relevant values are never presented.

While Toppino and Johnson (1973, Experiment I) found virtually no difference between the positive and negative (FF) conditions, this finding may be attributable to their use of distributed grid stimuli. These stimuli were generated on a 12-cell grid (four columns and three rows) by illuminating 1 cell (value) in each of the four columns (dimensions). Thus, the nonpresented stimulus values were always present and visible but simply unilluminated. This characteristic of grid stimuli may have reduced the difficulty of solution in Toppino and Johnson's (1973, Experiment I) positive (FF) condition because, if *S* attended to the unlit cells of the grid, the (FF) problem could be solved as a conjunctive (only the relevant values were never illuminated) by using an *affirmational strategy*. However, an affirmational strategy would have been totally inappropriate if conventional geometric stimuli had been employed because, with geometric stimuli, nonpresented values would be truly absent.

The present experiment employed inclusive disjunctive concepts and their complementary joint denial concepts in order to generate two conditions of category composition [(TT, TF, FT) and (FF)] crossed with two conditions of response label ("example" and "nonexample"). In addition, conventional geometric stimuli were employed instead of distributed grid stimuli. It was predicted that the use of these stimuli would substantially increase the difficulty of solution in the positive (FF) condition. The result should be an interaction between category composition and response label such that the positive label would yield superior performance in the (TT, TF, FT) conditions, while the negative label would yield superior performance in the (FF) conditions. Finally, each *S* was presented two successive problems to solve in order to investigate the stability of the interaction between category composition and response label.

[4] For a stimulus population defined on four three-valued dimensions, there are 9 TT instances, 18 TF instances, 18 FT instances and 36 FF instances. Thus, the (TT, TF, FT) category contains 45 stimuli and each relevant value appears on 27 stimuli. Therefore, the relative frequency of relevant values for the (TT, TF, FT) condition is .60. By similar computations, it can be shown that the relative frequency of relevant values is .50 for the (TF, FT) condition, .20 for (TT, FF) condition, and .00 for the (FF) condition.

Method

Su[illegible] design. Subjects were 50 introductory psych[illegible] [illegible]dents who earned extra class credit for exp[illegible]mental participation. Two *S*s were dismissed [illegible] to equipment malfunction. The remaining *S*[illegible] were randomly assigned by blocks of 4 to the four [illegible]xperimental conditions, generated by requiring *S* [illegible] use only examples or only nonexamples to solve either an inclusive disjunctive problem or a joint d[illegible]nial problem. In addition, each *S* was required to solve two succ[illegible] problems in his particular experiment[illegible]

[illegible]s. The stimulus population co[illegible] valued dimensions resulting [illegible] All *S*s were told that the [illegible] would vary along the dimensions o[illegible] [illegible]quare, or circle), arrow orientatio[illegible] ([illegible] up, sideward, or down), dot position (a[illegible] top, middle, or bottom), and border (dotted, da[illegible] or solid).

[illegible]timuli were mounted on 35-mm. slides and wer[illegible] projected onto a translucent screen by mea[illegible] of a Kodak Carousel 750 projector. Two resp[illegible] push buttons labeled "example" and "nonexa[illegible]" were located on a box situated in front of *S* [illegible]nd directly below the screen.

T[illegible]ee pairs of relevant values (solid border and dot in the middle, triangle and arrow pointing up, c[illegible] and dot at the bottom) were selected at random from among the 54 possible pairs. Each selected pair was relevant for one of the three problems which were prepared for each condition. Two of the three problems were presented to each *S* and the problems were assigned so that, in every condition, each of the three problems was presented to one third of the *S*s as the first problem and to another one third of the *S*s as the second problem. Sequencing restrictions used in the construction of these problems is described in detail in Experiment I of Toppino and Johnson (1973).

Procedure. The procedures of the present experiment were exactly the same as those detailed in Toppino and Johnson's Experiment I (1973). Briefly, *S* was given extensive rule training on his particular rule, followed by a test of rule acquisition to insure that he had achieved a maximally high level of rule learning prior to AI training. When *S* had correctly classified eight consecutive stimuli in the rule acquisition test, the nature of the AI task was thoroughly explained. A study–test procedure was employed, in which blocks of five training trials containing all positive or all negative stimuli were alternated with blocks of four test trials (without feedback) containing both positive and negative instances. Criterion was 12 consecutive correct responses accumulated over three or four test blocks.

Results

Analyses of trials and error data yielded virtually identical results. Therefore, only trials data are reported.

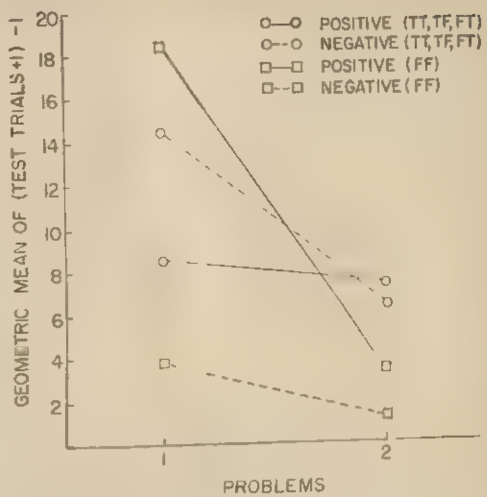


FIGURE 1. Mean test trials to criterion in terms of the antilogs of the means of $\log_e(x+1)$ minus one.

Test trials to criterion on the AI task were transformed into $x' = \log_e(x+1)$ in order to approximate homogeneity of variance and symmetry of distributions. The data are presented in Figure 1 in terms of the geometric means of (test trials to criterion $+1) - 1$ [i.e., antilogs of the means of $\log_e(x+1) - 1$].

Inspection of the data in Figure 1 suggests that the predicted two-way interaction between category composition and response label was obtained on Problem 1 but not on Problem 2, implying a three-way interaction. An overall analysis of variance (category composition, response label, pairs of relevant values, and problems) indicated that the overall two-way interaction between category composition and response label was marginally significant, F (1, 24) = 3.45, $p < .10$, and that the three-way interaction with problems approached significance, F (1, 24) = 2.81, $p \approx .11$. However, a more precise test of the experimental predictions was afforded by a separate analysis of variance for each problem.

On the initial problem, the predicted interaction was significant, F (1, 36) = 5.30, $p < .05$. The positive label yielded superior performance in the (TT, TF, FT) condition, while the negative label yielded superior performance in the (FF) condi-

tion. Neither category composition nor response label yielded a reliable main effect, both Fs $(1, 36) < 1.19$, $p > .20$.

Performance on the second problem indicated that practice eliminated the Response Label × Category Composition interaction, $F(1, 36) = .82$, $p > .20$, leaving only a main effect of category composition. The (FF) condition was solved more easily than the (TT, TF, FT) condition, $F(1, 36) = 7.77$, $p < .01$.

Discussion

The most important aspect of the present results was the interaction which occurred between response label and category composition on Problem 1. The nature of this interaction indicated that a positive label facilitates processing of (TT, TF, FT) instances, while a negative label facilitates processing of (FF) instances.

The hypothesis that predicted the obtained interaction represents an extension of an earlier hypothesis proposed by Bruner, Goodnow, and Austin (1956). According to these authors, preexperimental experience with conjunctive and affirmational concepts leads *S* to base solution attempts on stimulus elements which are common to positive instances (i.e., an affirmational strategy). The present hypothesis expands Bruner et al.'s notio[illegible] proposing that this same preexperimental [illegible]ould lead *S* to eliminate stimulus elemen[illegible]h appear in negative instances (i.e., a n[illegible]al strategy). Finally, if the effect of r[illegible]e label is based on preexperimental expe[illegible] it might be expected that the effect woul[illegible]ish with practice as observed in the pres[illegible]ults.

References

Bruner, J. S., Goodnow, J. J., & [illegible], G. A. *A study of thinking.* New York: [illegible] 1956.

Davidson, M. Positive versus ne[illegible]stances in concept identification proble[illegible]hed for logical complexity of solution proced[illegible] *Journal of Experimental Psychology,* 1969, 8[illegible]373.

Freibergs, V., & Tulving, E. T[illegible]ects of practice on utilization of informati[illegible]m positive and negative instances in conc[illegible]ntification. *Canadian Journal of Psycholo*[illegible]61, **15,** 101–106.

Haygood, R. C., & Bourne, L. E., Jr[illegible]ribute- and rule-learning aspects of conceptu[illegible]havior. *Psychological Review,* 1965, **72,** 175–1[illegible]

Hovland, C. I., & Weiss, W. Transm[illegible] of information concerning concepts throu[illegible]ositive and negative instances. *Journal of E*[illegible]*imental Psychology,* 1953, **45,** 175–182.

Toppino, T. C., & Johnson, P. J. [illegible]ts of category composition and response la[illegible]n attribute identification concept performan[illegible] *Journal of Experimental Psychology,* 1973, 10[illegible]–295.

(Received September 22, 1973)

Journal of E[illegible] Psychology
1974, Vol. 1[illegible] 1039–1045

STRE[illegible] OF AUDITORY STIMULUS–RESPONSE COMPATIBILITY AS A FUNCTION OF TASK COMPLEXITY[1]

JAMES CALLAN, DIANE KLISZ, AND OSCAR A. PARSONS[2]

Center [illegible] Alcohol Related Studies, University of Oklahoma Health Sciences Center

Ear response side and ear-hand components of lateral stimulus-response (S–R) compatibility were examined as a function of information processing [illegible] experiments. Experiment I investigated the compatibility [illegible] (two stimuli–one response) and Donders' *b* (two stimuli–[illegible] choice reaction time (RT) paradigms. A second experiment [illegible] effects in a Donders' *a* (simple) RT paradigm. Neither ear-[illegible] response location correspondence were evident in the *a* RT, [illegible] significant in the *b* and *c*. Ear hand correspondence was not [illegible] affected by task complexity but ear-response side correspond-[illegible]ntly increased from *c* to *b* and became more symmetrical. The [illegible] significantly faster to right ear stimulation in the *a* task but signi-[illegible]ly faster to left ear stimulation in tasks *b* and *c*. The results are discussed [illegible] terms of stage theory of information processing and in terms of their relevance for examination of functional asymmetry.

A [illegible] series of experiments has demonstra[illegible] the existence of a potent stereotypic response pattern (Simon, Hinrichs, & Craft, 1970). In a typical experimental paradigm *S*s respond to stimuli, which are presented to either the right or left side, by manual deflection of response keys located to the right and left of the body midline. The consistent finding has been that *S*s respond more rapidly when the (apparent) side of stimulus presentation and side of response execution correspond. Simon, Hinrichs, and Craft (1970) have identified ear-hand correspondence and ear-response location correspondence as two significant factors in this lateral stimulus-response (S–R) compatibility effect. In a choice (two stimuli–two response) reaction time (RT) task in which the content of a monaural tone (high or low) determined the responding hand, response side (as opposed to hand used) was identified as the most significant factor. Locus of the tone stimulus (right ear or left ear) was irrelevant in this task. When locus of a tone stimulus, which contained no information content (i.e., only one tone presented instead of two tones) determined the responding hand, both ear–response side and ear–hand correspondence effects contributed to the lateral S–R compatibility effect.

Simon and his colleagues have demonstrated this lateral S–R compatibility with both auditory and visual stimuli. They have manipulated the strength of the response pattern by (*a*) decreasing the directional information of the stimulus (Simon, Craft, & Small, 1971), (*b*) causing a conflict between the directional information in the stimulus content and in the stimulus location (Simon & Rudell, 1967), (*c*) introducing higher intensity stimulation (Simon, Craft, & Small, 1970), and (*d*) adding irrelevant directional stimuli in the same or different modalities (Craft & Simon, 1970; Simon & Craft, 1970; Simon, Craft, & Small, 1970; Simon & Small, 1969).

As an explanation for this phenomenon, Simon, Hinrichs, and Craft (1970) have postulated that it is the reflection of an "innate tendency [in humans] to respond toward the source of stimulation [p. 101]." However, the fact that this effect has been observed only in choice RT studies raises questions as to the generality of the above explanation. For example, is the compatibility effect also present in experimentally comparable, simple RT para-

[1] This research was supported in part by U. S. Public Health Services Grant 14702, National Institute of Mental Health. The authors express their appreciation to James Bertera for his assistance in the conduct of these experiments.

[2] Requests for reprints should be sent to Oscar A. Parsons, University of Oklahoma Health Sciences Center, P. O. Box 26901, Oklahoma City, Oklahoma 73190.

digms? Further, what stage or stages of information processing are involved in the compatibility effect? Is the effect primarily due to stimulus choice or response choice or both?

Information processing theory offers a conceptual framework within which the lateral S–R compatibility effect can be explored. According to this approach, a choice RT paradigm contains four information processing stages (Smith, 1968); (*a*) stimulus preprocessing, (*b*) stimulus categorization, (*c*) response selection, and (*d*) response initiation. Donders' (1969) three RT paradigms provide a systematic approach which enables inferences to be made concerning stages of information processing. Briefly, Donders' *a* is a simple RT paradigm (one stimulus, one response), Donders' *b* is a choice RT paradigm (two stimuli, two responses), and Donders' *c* is a go–no-go choice RT paradigm (two stimuli, one response). Donders' *b* compared with *c* allows the role of response choice to be examined while Donders' *c* compared with *a* enables the role of stimulus categorization to be investigated. In this article, two experiments are reported which examine both the ear–response side and ear–hand components of the lateral S–R compatibility effect as a function of information processing complexity within the framework of Donders' three RT paradigms. The results are pertinent to the following specific questions (which are addressed to both components of the compatibility effect): (*a*) Can the lateral S–R compatibility effect be replicated (Donders' *b*)? (*b*) Is the lateral S–R compatibility effect differentially affected by limiting response choice in a go–no-go paradigm (Donders' *c*)? (*c*) Is the effect present when both stimulus choice and response choice are eliminated as in a simple RT experiment (Donders' *a*)?

Experiment I

The purpose of Experiment I was to answer the first two questions posed above. Both choice (Donders' *b*) and go–no-go (Donders' *c*) RT paradigms were given to the same *S*s and the comp[illegible]ty effects were compared.

Method

Subjects. The *S*s were 16 femal[illegible] [illegible]6 male paid volunteers. All *S*s were right-hand[illegible] [illegible]d reported having normal hearing. The *S*s ra[illegible] in age from 18 to 31 yr. with a mean of 23

Apparatus. The *S*s were seated [illegible] [illegible]mfortable chair with arm rests on either side. A [illegible]onse panel was located directly in front of *S*'[illegible] [illegible]air. Two response keys were located on th[illegible] panel so that they could be reached without [illegible] the forearm from the arm rests. One res[illegible] key was located 15 cm. to the right, the other [illegible] 15 cm. to the left of *S*'s midline. The response [illegible]s were of the standard telegraph type, calibrate[illegible] [illegible] that the gap between contacts was maintained a[illegible] mm. and a pressure of 228 gm. was sufficient to m[illegible] contact. All responses were made by depressio[illegible] [illegible] the key with the index finger from a position i[illegible] which the finger rested lightly on the button to b[illegible] [illegible]epressed.

The stimuli were 200- and 500-Hz. tones generated by two Hewlett Packard Model [illegible]0 oscillators. The tones were presented throu[illegible] [illegible]atched pair of TDH 39 earphones which wer[illegible] [illegible]alibrated within 1 db. of each other. Output leve[illegible] the earphones was maintained at 90 db. (re .00[illegible] [illegible]b.) and was checked at regular intervals by a [illegible]librated sound level meter. The intensity rem[illegible]ed constant throughout the experiment. Tone [illegible]resentation and response time measurements w[illegible]re controlled by BRS 200 series digital pro[illegible]ramming equipment. Response times were recorde[illegible] in milliseconds on a Lehigh Valley 420-9 printout counter.

A green warning light located midway between the two response keys was presented 1 sec. prior to the onset of each tone, and there was a 4-sec. interval between tone onsets. Both the warning light and the tone remained on either until *S* made a response or until 1.5 sec. had elapsed from tone onset.

The *S* room was dimly lit by a small (6-w.) incandescent lamp located behind *S*. A 70-db. white noise background prevented disturbance from extraneous noises including the sound of the experimental equipment in an adjacent room.

Procedure. The *S*s were instructed to respond by depressing the key with their index finger. Each *S* performed under two RT conditions: a choice (Donders' *b*) and a go–no go (Donders' *c*). Half of the *S*s received the *b* RT condition first, and the others received the *c* RT condition first. Each *S* received eight practice trials before each condition.

In the *b* RT condition there were two blocks of trials. The *S*s responded on one block of trials with arms uncrossed, so that their left index finger was on the left key and their right index finger was on the right key. On the other block of trials *S*s responded with their arms crossed, so that their left index finger was on the right key and their right index finger was on the left key. Each block con-

TABLE 1

MEAN REACTION TIMES IN b AND c PARADIGMS

	Left hand				Right hand				
	Left side		Right side		Left side		Right side		$\bar{X}$
	$\bar{X}$	*SE*	$\bar{X}$	*SE*	$\bar{X}$	*SE*	$\bar{X}$	*SE*	
Choice reaction time (Donders' b)									
	377	12	447	16	390	18	442	17	414
	[illegible]8	14	421	15	453	19	382	13	429
	[illegible]18		434		422		412		421
Go–no-go reaction time (Donders' c)									
	318	11	339	13	323	13	355	14	334
	351	15	348	14	336	12	336	14	343
	335		344		329		346		338

Reaction time in milliseconds.

d of the same random sequence of 76 trials. ner a low pitched (200-Hz.) or a high pitched 0-Hz.) tone was presented monaurally on each al. Half of the *Ss* were told to respond with their right index finger when they heard the high-pitched tone and with their left index finger when they heard the low-pitched tone; the other half of the *Ss* were given the opposite hand–tone rule. Each *S* used the same hand–tone rule throughout the entire experiment. Within the b RT condition the crossed and uncrossed trial blocks were counterbalanced. In the crossed block, half of the *Ss* performed the task with the left arm over the right and the other half performed the task with the right arm over the left.

In the c RT condition there were four blocks of trials. Each block consisted of the same random sequence of tones that was used in the b RT condition. On each block of trials *S* was instructed to respond when he heard one tone, not to respond when he heard the other tone, and to place only one index finger on one response key. Every *S* performed one block of trials with each of the possible combinations of index finger and key positions: right-on-right, right-on-left, left-on-left, and left-on-right. The nonresponding arm was placed on the armrest alongside *S* out of the way of the responding hand. The order of presentation of the blocks of trials within the c RT condition was randomized.

Results

For each *S*, median RTs were calculated for each tone–ear combination. Means of the median RTs are presented in Table 1.

Planned comparison t tests were made on (a) order of presentation (b followed by c, c followed by b), (b) effect of hand–tone rule (high-right, low-left; low-right, high-left), and (c) effect of arm over in the crossed conditions (right-over-left, left-over-right). None of these tests were significant, and they were, therefore, eliminated from further analysis.

A five-way repeated measures analysis of variance (ANOVA) was conducted with condition, hand, ear, and side as repeated factors and sex as a between-*S* factor. The following main effects were significant: Sex, $F\ (1, 30) = 7.08$, $p < .05$, males responding faster than females (350 vs. 409 msec.); Condition, $F\ (1, 30) = 60.13$, $p < .001$, c responses were faster than b responses (338 vs. 421 msec.); Ear, $F\ (1, 30) = 18.81$, $p < .001$, left ear was faster than the right (374 vs. 386 msec.); and Side, $F\ (1, 30) = 4.78$, $p < .05$, left side was faster than the right (376 vs. 384 msec.).

The Ear × Side interaction $F\ (1, 30) = 109.71$, $p < .001$, successfully demonstrated the ear–response location correspondence effect and replicated the finding of Simon, Hinrichs, and Craft (1970) in that when stimulus and response sides corresponded, response times were significantly faster

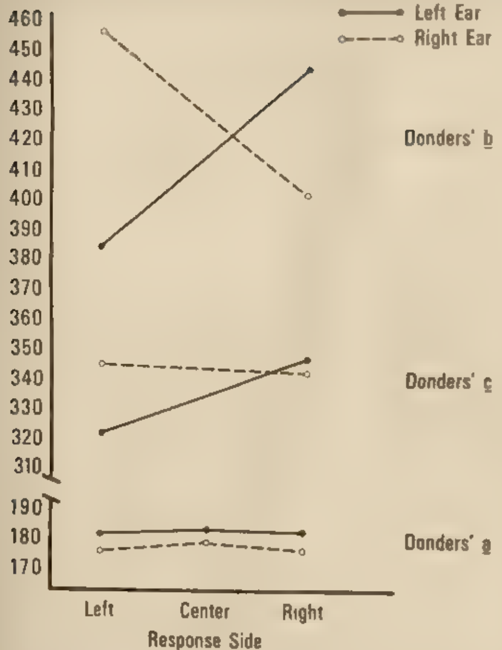


FIGURE 1. The Ear–Side correspondence effect in simple (Donder's *a*), two stimuli–one response (Donders' *c*), and two stimuli–two responses (Donders' *b*) RT conditions. (Right and left hands are collapsed.)

(Figure 1). The Condition × Ear × Side interaction, F (1, 30) = 67.43, $p < .001$, indicated that the ear–response location effect was significantly greater in the *b* than in the *c* condition (Figure 1). A second ANOVA was performed on the Ear × Side interaction within the *c* condition and the effect proved significant, F (1, 31) = 14.29, $p < .01$, although predominantly on the left side. Thus this correspondence effect, while stronger in the *b* RT, does appear in both conditions.

The Ear × Hand interaction was also significant, F (1, 30) = 18.53, $p < .01$, but the Condition × Ear × Hand was not (F = .04), indicating that the ear–hand correspondence effect, although present in both conditions, was not differentially affected by task complexity. The Ear × Hand interaction has another interesting feature; that is, the correpondence is entirely in the left hand. The means for right hand–right ear and right hand -left ear are both 377 msec. while the mean for the left hand–left ear was 370 msec. and for the left hand-right ear, 395 msec.

A comparison of Ear × Side and Ear × Hand correspondence effects are displayed in Table 2 for both *b* and *c* conditions. This table provides convenient comparison with the results of Simon, Hinrichs, and Craft (1970).

The only other significant effect was Sex × Condition × Ear, F (1, 30) = 10.84, $p < .01$; males had less difference between the right and left ear in the *c* condition than the females. This interaction may reflect differences in functional asymmetry between males and females as has been suggested by Kimura (1969). The failure of the Hand × Side interaction (F = .69) to achieve significance indicates that crossing the hands was not a significant factor. Simon, Hinrichs, and Craft (1970) found that uncrossed trials were significantly faster than crossed trials, but direct comparisons between their experiment and this cannot be made as the relative comfort of the arm positions is unknown.

EXPERIMENT II

The purpose of this experiment was to answer the third question, i.e., whether the lateral S–R compatibility effect would be found in a simple RT task (Donders' *a*)

TABLE 2

REACTION TIME AS A FUNCTION OF EAR–RESPONSE LOCATION CORRESPONDENCE AND EAR–HAND CORRESPONDENCE

Compatibility factor	Ear-hand correspondence	Ear-hand noncorrespondence	$\bar{X}$
Donders' *b* condition			
Ear-side correspondence	380	406	393
Ear-side noncorrespondence	450	450	450
$\bar{X}$	415	428	
Donders' *c* condition			
Ear-side correspondence	327	336	332
Ear-side noncorrespondence	338	353	346
$\bar{X}$	333	345	

Note. Reaction time in milliseconds.

in which bo[illegible] [illegible]timulus and response choice are elimi[illegible]ed. Simon (1967) examined ear prefer[illegible]e in a simple RT task with a centrally [illegible]ositioned response key and preferred [illegible] [illegible]nd responding. The possibility of an [illegible]teraction with response side was, there[illegible]re, not examined. In the present exp[illegible]ment, lateral and central response lo[illegible]ations are compared with right-ear, left-ear, and binaural stimul[illegible].

[illegible]

[illegible] male paid volunteers. All [illegible] as determined by self-report and [illegible] [illegible]rvation of writing habits. All *Ss* repo[illegible] normal hearing. Ages of the *Ss* range[illegible] 17 and 37 yr. of age with a mean of 26.

Ap[illegible]. The apparatus was basically the same [illegible] described in Experiment I with the addit[illegible] a third response key located on the *S*'s midli[illegible]dway between the two lateral response keys. [illegible]muli were 200-Hz. tones of the same intensity as in Experiment I. A green warning light locat[illegible] 15 cm. above the central response key preceded the stimulus tone by a randomly varying interval, either .5, 1.0, or 1.5 sec. The interstimulus interval was 5 sec. between tone onsets.

Procedure. Eight blocks of trials were presented. Six consisted of single-hand responses with each hand performing in turn on each response key. Two blocks of trials consisted of two-hand responding (crossed and uncrossed) in conjunction with another experiment. During two-hand responding trials, the *S*'s task was to respond with both index fingers simultaneously while *E* recorded RT in milliseconds to the first key depressed. Half of the *Ss* received the two-hand block followed by the single-hand block, and the other half of the *Ss* received the trial blocks in the reverse order. The order of single-hand trial block presentation was randomized across all *Ss*.

The *Ss* were given 21 practice trials randomly presented so that there were 7 each in the right ear, the left ear, and binaurally. Practice responses were on a randomly chosen single-hand response key. Each trial block consisted of 63 randomly presented stimuli, 21 each to the right, left, and both ears. In between trial blocks, *S* was told that he had performed very well and then told which hand–response key condition would follow. The *Ss* were exhorted at the beginning of the experiment and between each trial block to respond as rapidly as possible. Further, they were told that the tone would appear sometimes in the right ear, sometimes in the left, and sometimes in both. Regardless of where the tone appeared, they were to respond as rapidly as possible with the assigned hand–key condition.

Results

Median RTs were calculated for each of the ear presentation locations (right, left, binaural) within each of the trial blocks. Planned comparison *t* tests of single-hand vs. two-hand response trial blocks revealed no significant differences. Further *t* tests of single-hand responding before vs. single-hand responding after two-handed response trials revealed no significant differences. The two-hand responses were omitted from further analyses. A three-way (Ear × Hand × Side) repeated measures ANOVA was then performed on the median scores of the single-hand response trial blocks. Table 3 presents the median scores. Of the seven possible main and interaction effects, only two were significant: Ear, $F(2, 58) = 16.22$, $p < .01$, and Hand × Side, $F(2, 58) = 3.94$, $p < .05$. The right-ear and binaural RTs did not differ, but both were significantly faster than the left-ear

TABLE 3

Mean Reaction Times in the *a* Paradigm

Ear	Left hand: Left side $\bar{X}$	SE	Left hand: Center $\bar{X}$	SE	Left hand: Right side $\bar{X}$	SE	Right hand: Left side $\bar{X}$	SE	Right hand: Center $\bar{X}$	SE	Right hand: Right side $\bar{X}$	SE	$\bar{X}$
Left	186	8	181	6	182	8	176	6	183	7	181	6	181
Binaural	181	8	186	6	173	7	172	6	178	7	177	6	176
Right	180	7	177	6	175	7	172	7	181	6	177	6	177
$\bar{X}$	182		178		177		173		181		178		178

Note. Reaction time in milliseconds.

responses: left-ear mean = 181 msec.; binaural mean = 176 msec.; right-ear mean = 177 msec. Figure 1 presents the means of right- and left-ear responses on the three response keys for comparison with the results of Experiment I.

The significant Hand × Side interaction reflected the fact that the right hand was faster when operating on the left-side response key; similarly, the left hand was faster on the right side. This effect was not present in Experiment I and may represent a chance finding.

Discussion

Lateral stimulus–response compatibility effects. Those results which are relevant to the specific experimental questions are as follows: (*a*) The lateral S-R compatibility effect described by Simon, Hinrichs, and Craft (1970) in the Donders' *b* paradigm is quite strong and reproducible. (*b*) The ear-response location component of the lateral S–R compatibility effect is present under the Donders' *c* RT condition but is significantly less than in the Donders' *b* condition. At the same time the symmetry of the effect increases from *c* to *b*. The ear–hand correspondence effect is consistently present in both *b* and *c* conditions but operates almost exclusively in the left hand. (*c*) There is no S-R compatibility effect in a simple RT situation, Donders' *a*.

The presence of the Ear × Side interaction in the Donders' *c* paradigm and its absence in the Donders' *a* paradigm suggests that the stimulus categorization stage, but not stimulus preprocessing or response initiation, is important with respect to the ear–response location correspondence effect. When the additional stage of response selection is added in the Donders' *b* paradigm, this correspondence effect significantly increases. Response choice appears to be an important determinant of the effect. The ear–hand correspondence is also present in both *b* and *c* paradigms, but not differentially so, and is absent in Donders' *a*. Stimulus categorization, therefore, appears to be the most important stage for the ear–hand correspondence effect. It is also important to note that both left-ear–hand and left-ear–response location correspondence effects are stronger than those for the right. Further, while ear–response location correspondence becomes more symmetrical as the task complexity increases from *c* to *b*, ear–hand correspondence is almost entirely a left-hand effect.

Ear effects. The significantly faster response to right-ear stimulation in the simple RT (Experiment II) is consistent with the finding of Simon (1967). However, it is interesting to note that in the choice RT conditions (Experiment I), responses to left-ear stimulation were faster than to right-ear stimulation, a reversal of the relationship in the simple task. This reversal may have been due in part to the additional stimulus processing required in the stimulus categorization stage. Kinsbourne (1970) has suggested that the strong right-ear bias in the simple RT may reflect an attentional preference for the left hemisphere as a consequence of *S*s concurrent verbal thinking processes, i.e., the hemisphere in use displays a lower attentional threshold. The results from these experiments suggest another explanation within the framework of Kinsbourne's attentional hypothesis. The more complex *b* and *c* choice tasks appeared to require more concentrated concurrent verbal activity as *S* had to remember the hand–tone rule, e.g., "high tone–right hand." The significant left-ear effect in these situations might have been due to the preoccupation of the left hemisphere with verbal rehearsal. In fact, some *S*s reported rehearsing the rule over and over to themselves during the task. Systematic control of this variable, however, was not attempted in this experiment.

A second explanation for the ear effects in the *b* and *c* conditions may be that the right hemisphere is more effective in tonal discrimination than the left hemisphere. The work of Kimura (1967) provides support for this hypothesis. This would explain the facilitation of left-ear stimulus processing in those conditions in which high and low pure tones must be discriminated before responding. Perhaps both left hemisphere interference and right hemisphere facilitation are operating to produce the results obtained in the choice RT conditions.

In conclusion, the data from these experiments suggest that information processing complexity, as exemplified in the additive model of Donders, must be taken into account when making inferences about the generality of the lateral S–R compatibility effect. These data further suggest that lateral asymmetries in RT performance are affected by the manipulation of information processing complexity.

Craft, J. L., [illegible] cessing symbolic informati[illegible] y: Interference from an i[illegible] te. *Journal of Experi[illegible]* [illegible] 415–420.

Donder[illegible] tal processes. (Tra[illegible] ca, 1969, 30, 412–431

Kimura, D[illegible] the brain in [illegible]

Kim[illegible] ght vi[illegible] ology, 1969, 23, [illegible]

Kinsbourne, M. Th[illegible] basis of lateral asymmetries in [illegible] *Psychologica*, 1970, 33, 193–20[illegible]

Simon, [illegible] in a simple reaction-time task [illegible] *xperimental Psychology*, 1967, 7[illegible]

Simon, J. R., & C[illegible] J. L. Effects of an irrelevant auditory s[illegible] n visual choice reaction time. *Journal o[illegible]rimental Psychology*, 1970, **86**, 272–274

Simon, J. R., Craft, J. L., & Small, A. M., Jr. Manipula[illegible]ng the strength of a stereotype: Interference effects in an auditory information-processing task. *Journal of Experimental Psychology*, 1970, **86**, 63–68.

Simon, J. R., Craft, J. L., & Small, A. M., Jr. Reactions toward the apparent source of an auditory stimulus. *Journal of Experimental Psychology*, 1971, **89**, 203–206.

Simon, J. R., Hinrichs, J. V., & Craft, J. L. Auditory S–R compatibility: Reaction time as a function of ear-hand correspondence and ear-response location correspondence. *Journal of Experimental Psychology*, 1970, **86**, 97–102.

Simon, J. R., & Rudell, A. P. Auditory S–R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 1967, **51**, 300–304.

Simon, J. R., & Small, A. M., Jr. Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, 1969, **53**, 433–435.

Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

(Received September 24, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1046–1052

ESTIMATION OF WORD FREQUENCY IN CONTINUOUS AND DISCRETE TASKS [1]

IAN BEGG [2]

McMaster University, Hamilton, Ontario, Canada

Long lists of concrete and abstract nouns which appeared varying numbers of times were presented to two groups of *S*s. One group judged how many times each word occurred during presentation (online), and again after the whole list had been presented (delayed). The second group only judged the words after the list had been presented (discrete). All *S*s subsequently recalled the words. Frequency estimates were above true frequency for relatively low frequencies, and below for relatively high frequencies in both the delayed and discrete judgments, but more so in the latter. Additionally, the interaction was more pronounced in abstract than concrete nouns. Recall was a positive function of concreteness, true frequency, and subjective frequency. The results were discussed in terms of multiple-trace and multiple-process views of how frequency is represented in memory.

This paper addresses the question of how *S*s estimate the number of times a word has occurred in a given situation. According to Howell (1973a), there are four major hypotheses concerning the representation of event frequency in memory. The first, the trace-strength hypothesis, maintains that each presentation of an event causes an increment in the strength of the memory trace corresponding to the event, and that frequency estimation is effected by reading out current strength values. The second, the multiple-trace hypothesis, maintains that each presentation of an event produces a new trace (Hintzman & Block, 1971), or a new "list marker" (Anderson & Bower, 1972), and that frequency estimation consists of counting the number of traces or list markers. What the two hypotheses have in common is that there is no unique frequency representation process. Rather, frequency estimation consists of inferring numerical judgments from a general memory representation. What differentiates the hypotheses is that the trace-strength hypothesis has no mechanism for separately retaining information about the individual contexts in which each presentation of an event occurs. Since *S*s can retrieve information about the duration (Hintzman, 1970) and modality (Hintzman, Block, & Summers, 1973) of specific occurrences of events, the multiple-trace hypothesis is more useful than the trace-strength hypothesis, which will not be discussed further here.

Two further hypotheses can be derived from the proposition that the situational frequency of occurrence of an item is specifically represented in memory, and is not simply inferred from the general memory representation. One such hypothesis, the numerical-inference hypothesis, is that *S*s simply count the number of times items have occurred, and give the current count as the frequency judgment. The second hypothesis, the frequency-attribute hypothesis, is that an item is remembered as a set of attributes, one of which increments with situational frequency; estimates of event frequency could thus be made by reading out the frequency attribute (Underwood, 1969a). Since the numerical-inference hypothesis can be subsumed by the frequency-attribute hypothesis, they will both be referred to as multiple-process hypotheses, since they posit that frequency information is specifically represented in memory, rather than as an incidental component of the general memory representation of an event.

It is difficult, if not impossible, to differentiate experimentally between a multiple-

[1] This research was supported by National Research Council of Canada Grant A8122.

[2] Requests for reprints should be sent to Ian Begg, Department of Psychology, McMaster University, Hamilton, Ontario, Canada L8S 4K1.

trace and a multiple-process hypothesis. Since memory can be conceived of as being characterized by different levels of processing (Craik & Lockhart, 1972), words processed for memory could simply be processed to a deeper level than words processed for frequency, so that differences in memorability between words processed under the two conditions could either reflect different processes, as [illegible]ied by Underwood ([illegible]), or reflect the fact that words processed under frequency instructions were not processed as deeply as words processed under memory instructions. Howell ([illegible]) found that words processed under frequency instructions were not as well recalled as those processed under memory instructions, but that frequency judgments were not affected by the set under which the events were processed. Howell concluded that the memory difference reflected a different process, although it is also possible to interpret the results in terms of different levels of processing.

If either the multiple-trace or the multiple-process hypothesis is to provide a reasonable account of frequency estimation, there are certain basic facts that must be taken into account. First, if *S*s are presented with a list of words that appear varying numbers of times, and judge on every trial the number of times a particular word has occurred, average judgments correspond closely to actual frequency (Begg & Rowe, 1972, Experiments II & III). Either hypothesis could handle the finding easily. For the Anderson and Bower (1972) multiple-trace model, the mean number of list markers available for an item presented n times in a list is $n\ \alpha$, where α is the probability of establishing a list marker at the time of presentation. The mean number of list markers has an upper bound defined by true frequency (n) and a lower bound of zero. Frequency estimates are provided by a transformation of the number of list markers. The results reported by Begg and Rowe could be accounted for by making the transformation linear (i.e., $Fn = a + bn\alpha$), in which case $a = 0$, and $b\alpha = 1$. The multiple-process hypothesis could handle the result by assuming that the frequency attribute, on the average, increments by a unit amount for every unit presented and is, at least during acquisition, generally retrievable.

The second general finding to be accounted for is that if frequency judgments are not made until after the list has been presented, the judgments generally are higher than relatively low true frequencies, but below relatively high frequencies (e.g., Peterson & Beach, 1967). There are several points to note about the interaction between true and judged frequency. First, the interaction obtains when *S*s are previously instructed either that the task will require frequency judgments (e.g., Howell, 1973b; Underwood & Freund, 1970) or to remember the words (e.g., Hintzman, 1969; Howell, 1973b). Second, the interaction is sensitive to the frequencies used in the specific task, since it occurs with as few as 2 presentations representing the highest frequency level (Hintzman et al., 1973), or as many as 10 (Howell, 1973b). That is, the interaction is a function of relative, rather than absolute, frequency. Third, a retention interval makes the interaction more pronounced, both with respect to overestimation of low and underestimation of high relative frequencies (Underwood, Zimmerman, & Freund, 1971).

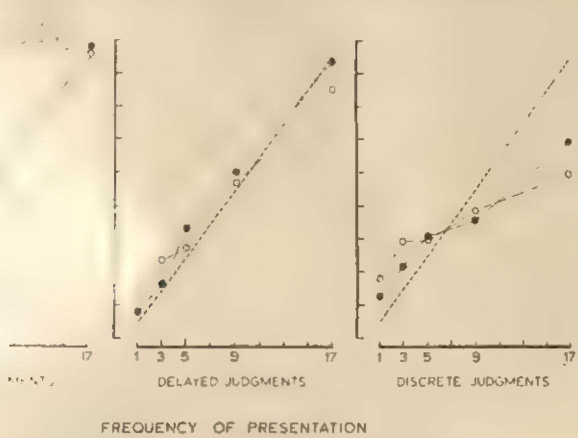
Since true and judged frequency interact when judgments are made after presentation of a list (discrete judgments), but not during presentation of the list (online judgments), the interaction could reflect one of two factors, or a combination of the two. Online and discrete judgments differ both with respect to the presence of a delay interval interposed between the last presentation of an item and the subsequent estimation in the latter case, and with respect to the presence of overt running judgments in the former. In the present experiment, *S*s who made online judgments were also asked to judge the words at the end of the task (delayed judgments). Comparisons between the two sets of judgments reflect retention effects, while comparisons between delayed and discrete judgments reflect the effects of the overt running judgments made in the former condition.

The facts that true and judged frequency interact, that the interaction is relative to the range of true frequencies used, and that a retention interval accentuates the interaction, must be accounted for by the hypotheses, either by changes in the memory representation over time, or by variables affected by the presence of overt judgments during acquisition. First, consider the effects of a retention interval on the representations from which frequency estimates are inferred. According to Anderson and Bower (1972), the effect of a retention interval is to reduce the availability of individual list markers, so that the best fit line relating true and judged frequency would become less steep over time. The problem is that the model has no explicit provision for raising the intercept of the line as a function of forgetting, and so does not account for (*a*) the overestimation of low-frequency words, (*b*) the fact that the interaction is relative to the range of experimental frequencies, or (*c*) the exaggeration of the degree of overestimation of low frequencies with longer delay intervals. However, the problem is not serious. If *S* compares his subjective mean number of list markers available across the full range to his subjective estimate of the mean number of occurrences of words, the difference between the two could simply be added to all estimates derived from the number of list markers available. That is, the effect of a retention interval is to move all judgments closer to the task-defined mean frequency. For Underwood's (1969a) multiple-process model, no specific statement is made about the effects of a retention interval on the frequency attribute. A reasonable result of deterioration of the frequency attribute, however, is that the attribute will become assimilated to the mean situational frequency of occurrence of the words. Thus, judgments would generally become closer to the task mean over increasing retention intervals. Consequently, both models of frequency estimation can account for a frequency interaction, which is a function of relative frequency, and which becomes more pronounced with time. The forms of both models as developed here would also predict that the overall average of responses would not differ as a function of retention interval, as found by Underwood et al. (1971).

Recall that, in addition to a retention interval, online and discrete judgments differ in that *S*s in the former case make running judgments during the presentation of the words. Whether such estimation requires closer attention, or provides retrieval practice, or whatever, both hypotheses, as interpreted above, would yield the prediction that the effects of such judgments would be effectively opposite to the effects of forgetting. That is, removing the running estimates should cause judgments to become closer to the task mean.

In the present experiment, concrete and abstract words were presented to *S*s who either judged the frequency of occurrence of each word as it occurred (online) and then again after the list had been presented (delayed), or only judged frequency after the presentation of the entire list (discrete). The expectation is that the online judgments will be relatively accurate, the delayed judgments will show an interaction between true and judged frequency, and that the interaction will be most pronounced in the discrete judgments. The effects of concreteness on frequency estimation are not very pronounced in online judgments (Begg & Rowe, 1972). Since Anderson and Bower (1972) conceive of tagging with list markers as essentially a paired-associate task, abstract words should show a more pronounced frequency interaction than concrete words with both delayed and discrete judgments, because (*a*) concrete words are known to be better stimuli than abstract words in paired-associate learning (e.g., Paivio, 1969), and (*b*) concrete words are better retained than abstract words over long periods of time, particularly if the recall task is associative (Begg & Robertson, 1973). A frequency attribute, if it is conceived of as a paired-associate of the word, would yield the same predictions as the Anderson and Bower model.

The *S*s in the present experiment were also asked to recall the words after they



1. Online, delayed, and discrete judgments of frequency, as a function of concreteness and the frequency of presentation.

had made their terminal judgments. Recall was expected to be a positive function of concreteness (e.g., Begg & Robertson, 1973), true frequency, and subjective frequency (e.g., Underwood et al., 1971).

Method

Subjects. Twenty volunteers from McMaster University were paid $2/hr for their participation. Half of the *Ss*, determined randomly, judged the frequency of each word as it occurred, and half made their first judgments after the entire list had been presented.

Materials and procedure. Fifty concrete and 50 abstract nouns were chosen from the Paivio, Yuille, and Madigan (1968) norms, with mean *I* ratings of 6.39 and 3.04. All words were A or AA in frequency of occurrence. Ten concrete and 10 abstract nouns were printed on 17 cards, 10 were printed on 9 cards, and another 10 were printed on each of 5, 3, and 1 card. The 700 cards thus produced were arranged in 28 sets of 25 cards, with successive occurrences of the words from each frequency level being placed in consecutive piles. The pile in which the initial presentation of an item occurred was as random as possible, given the constraint that there were enough adjacent piles for the later presentations of an item, and that an equal number of concrete and abstract nouns from any given frequency level started in the same pile. After the sorting, each pile was thoroughly shuffled, and the blocks were placed together to form one deck of 700 cards. On the average, then, 25 items separated successive presentations of a particular item.

The task was presented to each *S* in three stages. In the presentation stage, *S* went through the entire list of 700 cards at his own rate, reading each word aloud. The *Ss* in the continuous-presentation condition estimated the frequency of occurrence of each word in the list, while *Ss* in the discrete-presentation condition made no overt judgments, although they had been told that their task was to remember how many times each word occurred. In the second stage of the task, the judgment stage, all *Ss* were presented with lists containing the 100 words from the initial list. Six different random orders were used. The *Ss* were instructed to write down next to each word a number representing how many times they thought the word had occurred in the initial list. In the final stage of the task, the recall stage, all *Ss* were asked to write down as many of the words as they could remember, in any order. The whole task took about 75 min.

Results and Discussion

Frequency estimates. Since any departure from accuracy in judging the frequency of once-presented items would necessarily be above true frequency, median rather than mean judgments were analyzed. For *Ss* in the continuous-judgment task, two sets of medians were calculated, one for the online judgments made during presentation and one for the delayed judgments. Each set of medians had 10 members, 1 for the 10 concrete and 1 for the 10 abstract nouns at each of the five frequency levels. For the online judgments, only responses made to the last presentation of an item were scored. The medians for the delayed judgments and for the discrete judgments included all nouns in the final judgment list.

The mean medians, averaged across *S*s, are presented in Figure 1. For each of the three types of judgments, results are plotted separately for concrete and abstract nouns, and a perfect-accuracy line is included for reference purposes. There are several points to note in the figure. First, the online judgments made in the continuous task did not depart much from true frequency, as found by Begg and Rowe (1972), and there were no apparent effects of concreteness. However, when the same *S*s repeated their judgments in the delayed condition, the judgments showed an interaction between true and judged frequency, and the interaction was more pronounced for abstract than concrete nouns. The *S*s in the discrete task showed a more exaggerated interaction than *S*s in the delayed task, and the interaction was again most pronounced with abstract nouns. It is also perhaps of note that the means of the 10 medians in each judgment condition (7.1, 7.7, and 6.7 for the online, delayed, and discrete judgments, respectively) did not depart much from the true mean frequency of presentation, which was 7, suggesting that *S*s have some idea of how many presentations actually occurred, and adjust their responses accordingly.

The reliability of the results was determined by two analyses of variance. The first analysis compared the online and delayed judgments made in the continuous task, to assess the effects of a delay interval on the judgments. Thus, time of judgment (online or delayed), concreteness (abstract or concrete), and frequency of presentation were all repeated factors. Overall, judgments increased with frequency, F (4, 36) = 53.3, $p < .01$. More importantly, frequency and the time of judgment interacted, F (4, 36) = 7.44, $p < .01$, confirming the reliability of the tendency to overestimate low- and underestimate high-frequency nouns in the delayed judgments. Further, the three-way interaction, F (4, 36) = 2.65, $p < .05$, reflected the fact that the pattern found in the delayed condition was more pronounced with the abstract than concrete nouns. In fact, there is only a hint of the interaction in the delayed judgments of concrete nouns. The results can be seen in the two left-hand panels of Figure 1.

The second analysis compared the two sets of judgments made after all the words had been presented. Differences between the delayed judgments and the discrete judgments can be attributed to the running estimates made by the *S*s in the delayed task. The results were analyzed by 2 × 2 × 5 analysis of variance, with the type of task as an independent factor, and concreteness and frequency as repeated factors. As in the previous analysis, judgments increased with frequency, F (4, 72) = 67.3, $p < .01$. The task variable interacted with frequency, F (4, 72) = 6.18, $p < .01$, since the interaction between true and judged frequencies was more pronounced in the discrete than the delayed task. Finally, there was a significant Concreteness × Frequency interaction, F (4, 72) = 4.22, $p < .01$, since the interaction between true and judged frequencies was more pronounced with abstract than concrete nouns in both tasks, as seen in the two right-hand panels of Figure 1.

Overall, then, the interaction between judged and true frequency is absent in online judgments, is present in both delayed and discrete judgments, and is especially pronounced in the latter. Further, the interaction is more pronounced when abstract rather than concrete words are judged. The results conform quite well to the expectations of either a multiple-trace model that posits a comparison process between the number of list markers available and the number of presentations made, or to a frequency-attribute model that assumes the frequency attribute becomes assimilated into the task-defined mean frequency as it deteriorates.

If the individual judgments become assimilated to situational frequency, then it should be the case that the amount of variability around the mean decreases as the interaction between true and judged frequency becomes more pronounced. From online to delayed judgments, the mean standard deviation decreased from 6.39 to 5.89, F (1, 9) = 7.27, $p < .025$, for the

abstract nouns. The decrease was not found for [illegible] nouns, which did not show as pronounced an interaction as the abstract [illegible]. [illegible] the mean standard deviation in the delayed judgments was 6.[illegible] compared to [illegible] for the discrete judgments. A t test performed on $\log_{10}$ transformations of the standard deviations showed the [illegible] reliable, t (18) = 14.0[illegible] [illegible] general [illegible] the judged situa[illegible].

Recall data. The [illegible]er of words recalled by *S*s in [illegible] was analyzed by a 2 × 2 × [illegible] variance with task (cont[illegible]) as the independent factor, [illegible]teness and frequency as repeated [illegible]. Recall increased with frequency, F ([illegible], 72) = 10.3, $p < .01$, with respective mean recalls of 1.95, 2.40, 2.68, 3.28, and [illegible].88, each out of a possible 10. Concreteness was also positively related to recall, F ([illegible], 18) = 15.5, $p < .01$, with 16.8 and 11.5 concrete and abstract words recalled out of a possible 50. The Concreteness × Task interaction was not reliable ($F < 1$), although in absolute terms, the concreteness effect was larger in the discrete task (16.5 vs. 9.4) than in the continuous task (17.1 vs. 13.7).

In order to determine whether apparent frequency affected recall, the frequency judgments made to the words that were later recalled were compared to the judgments of words not recalled. Because of a number of empty cells, the data were averaged over low frequency (1, 3, and 5), high frequency (9 and 17), and concreteness. In the discrete task, the mean judgment of recalled words was higher than the judgment of forgotten words, both at low frequency, 6.23 vs. 5.12, F (1, 9) = 6.70, $p < .05$, and at high frequency, 11.07 vs. 8.92, F (1, 9) = 15.0, $p < .01$. The same was true in the continuous task, both at low frequency, 5.92 vs. 4.70, F (1, 9) = 7.22, $p < .05$, and at high frequency, 13.49 vs. 11.89, F (1, 9) = 5.83, $p < .05$. Thus, recalled words were consistently higher in judged frequency of occurrence than forgotten words.

As a second test of the hypothesis that subjective frequency is related to recall, recall was conditionalized on whether the final frequency judgments were above or below each *S*'s median judgment at each frequency level. For each *S*, the results were averaged over concreteness, and once-presented items were not considered, since for several *S*s there were no judgments below the median. The proportions of words recalled, given that the final frequency judgment was above or below the median, were analyzed by a 2 × 2 × 4 analysis of variance, with task (continuous or discrete) as an independent factor, and level of final judgment (above or below the median) and frequency (3, 5, 9, or 17) as repeated factors. Both subjective frequency, F (1, 18) = 13.5, $p < .05$, and actual frequency, F (3, 54) = 3.50, $p < .05$, were positively related to recall. Thus, items initially judged above the median were more likely to be recalled than items judged below the median, with respective proportions of .356 and .250, and recall increased over the four frequency levels included in the analysis, with proportions of .240, .280, .317, and .375. No other effects were reliable.

Thus recall is a positive function of both subjective and actual frequency of occurrence, whether subjective frequency is defined by average ratings of recalled words, or by recall or words partitioned according to the level of the frequency judgments. This extends the finding of Underwood et al. (1971).

The general conclusion of the paper is that any reasonable account of frequency estimation must include some process that allows judgments to become closer to the task mean, whether that process is an additive bias (multiple trace), or assimilation of a frequency attribute into that mean. In his review, Howell (1973a) did not consider the nature of the interaction between true and judged frequency, or the fact that the interaction is a function of relative, rather than absolute frequency. However, his conclusion that it is not possible to rule out the alternative hypotheses is entirely correct.

The present results also have implications for verbal discrimination learning, a task in which concrete items lead to better performance than abstract items even on the first trial (e.g., Rowe & Paivio, 1971). Ekstrand, Wallace, and Underwood (1966) offered an account of verbal discrimination in which situational frequency was viewed as the dominant attribute affecting performance. Such an account cannot easily explain imagery effects, since abstract words actually have higher subjective frequencies than concrete words for low frequencies if the judgments are delayed, and the judgments are not different if they are made at the time of presentation of the words. Thus, although situational frequency is a powerful explanatory construct with respect to verbal discrimination learning, additional constructs must be invoked to explain the effects of imagery in the task.

REFERENCES

Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, **79**, 97–123.

Begg, I., & Robertson, R. Imagery and long-term retention. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 689-700.

Begg, I., & Rowe, E. J. Continuous judgments of word frequency and familiarity. *Journal of Experimental Psychology*, 1972, **95**, 48–54.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 671–684.

Ekstrand, B. R., Wallace, W. D., & Underwood, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

Hintzman, D. L. Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Ps*[illegible] 1969, **80**, 139–145.

Hintzman, D. L. Effects of repeti[illegible]d exposure duration on memory. *Journ*[illegible]*perimental Psychology*, 1970, **83**, 435–44[illegible]

Hintzman, D. L., & Block, R. A. [illegible]tition and memory: Evidence for a multip[illegible] hypothesis. *Journal of Experimental Psy*[illegible], 1971, **88**, 297–306.

Hintzman, D. L., Block, R. A., & S[illegible]ers, J. J. Modality tags and memory for rep[illegible]ns: Locus of the spacing effect. *Journal of V*[illegible]*l Learning and Verbal Behavior*, 1973, **12**, 229[illegible]

Howell, W. C. Representation of [illegible]quency in memory. *Psychological Bulletin*, 19[illegible] 80, 44–53. (a)

Howell, W. C. Storage of events a[illegible] event frequencies: A comparison of two p[illegible]digms in memory. *Journal of Experimental* [illegible]*chology*, 1973, **98**, 260–263. (b)

Paivio, A. Mental imagery in associat[illegible] learning and memory. *Psychological Review*, [illegible]69, **76**, 241–263.

Paivio, A., Yuille, J. C., & Madig[illegible] S. A. Concreteness, imagery, and meaningful[illegible] values for 925 nouns. *Journal of Experim*[illegible]*l Psychology, Monograph Supplement*, 1968, 7[illegible], Pt. 2).

Peterson, L. R., & Beach, C. R. Ma[illegible] as an intuitive statistician. *Psychological Rev*[illegible], 1967, **68**, 29–46.

Rowe, E. J., & Paivio, A. Word freque[illegible] and imagery effects in verbal discrimination l[illegible]rning. *Journal of Experimental Psychology*, 1971, **88**, 319–326.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–574. (a)

Underwood, B. J. Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 83[illegible]94. (b)

Underwood, B. J., & Freund, J. S. Relative frequency judgments and verbal discrimination learning. *Journal of Experimental Psychology*, 1970, **83**, 279–285.

Underwood, B. J., Zimmerman, J., & Freund, J. S. Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology*, 1971, **87**, 149–162.

(Received September 29, 1973)

Journal of Exp[illegible]ntal Psychology
1974, Vol. 1[illegible] 1053-1060

HYPOTHESIS BEHAVIOR AS A FUNCTION OF AMOUNT OF PRETRAINING[1]

RAYMOND M. WHITE, JR.[2]

Miami University

Two ex[illegible]iments were performed to determine the effects of conceptual pretrainin[illegible] the hypothesis behavior exhibited by human *S*s on simultaneous discrim[illegible]n problems that included blank trials. For Experiment I, 80 *S*s [illegible] 1, 2, 3, or 4 simultaneous discrimination problems prior to four blank t[illegible] problems. Experiment II was defined by the factorial combination of [illegible], 4, or 6 pretraining problems and two sets of instructions, 128 *S*s being [illegible]en routine instructions or directions to consider a single hypothesis at a time. With training, *S*s were more likely to maintain their previous hy[illegible]esis after positive outcomes and to alter their hypothesis after negative [illegible]s. The conclusion was that the uniprocess hypothesis models, such [illegible]ne proposed in 1966 by Levine, do not accurately describe the behavior [illegible]trained *S*s, and that a multiprocess model of conceptual learning is [illegible]ted.

Many, if not all, theories of concept identification (CI) ascribe a single process to the problem solver in his attempt to identify the relevant attributes of a conceptual problem (e.g., Levine, 1966; Restle, 1962). However, recently White (1972a, 1972b) has discussed the possibility that *S* retains a number of means for solving a problem: Which of the approaches will be adopted in attempting to solve a problem depends on a number of factors such as the type and amount of training *S* has received.

Support for White's (1972a, 1972b) multiprocess view of CI was indirect. The *S*s in his experiments were given one of two types of pretraining problems. The performance of *S*s who were given systematic training (i.e., a routine conceptual problem) was unaffected by the length of postfeedback intervals after correct responses—indicating the absence of problem-solving activity during that interval—but strong effects due to the duration of the postfeedback interval after errors were observed. These results were as predicted by an error-trial-only theory of CI (e.g., Bower & Trabasso, 1964). On the other hand, the performance of *S*s who received unsystematic training, consisting of a problem whereby *S* could reach the solution criterion by memorizing the placement of each stimulus pattern, was affected equally by the length of intervals after both correct and incorrect responses. The latter result is consistent with the predictions of an all-trials, response-independent theory of CI (e.g., Bourne & Restle, 1959). Furthermore, as the amount of pretraining was increased, the effects of the length of postfeedback intervals were more consistent with the predicted outcomes of the opposing theories: Systematic training provided performance orderings as predicted by error-trial-only CI models, and unsystematic training produced orderings as predicted by all-trials CI models.

Levine (1966, 1970) has provided a more direct way of assessing the processes underlying performance in a CI problem and a theory to account for behavior in such problems. Levine's method is to provide certain trials during which *S* is given feedback concerning the correctness of his response (hereafter termed outcome trials, OTs). Following an OT, typically, are a series of trials for which no feedback is given (hereafter termed blank trials, BTs).

---

[1] The research reported herein was supported in part by the Faculty Research Committee at Miami University. The author wishes to express his sincere appreciation to Ed Coleman, Mike Dowdle, Kathy Nelson, Mark MacCormick, and especially to Mary Whalen for their assistance in preparing and conducting the experiments.

[2] Requests for reprints should be sent to Raymond M. White, Jr., Department of Psychology, Miami University, Oxford, Ohio 45056.

By observing the responses *S* makes during the BTs, the hypothesis *S* is entertaining can be determined. For example, in a simultaneous discrimination problem, if *S* always selects the stimulus appearing on the left side of the display as containing the correct attribute, it is inferred that he is hypothesizing that "leftness" defines problem solution. Through such analyses, it can be determined on what occasions *S* changes hypotheses and the manner in which the hypotheses are changed.

Levine (1966, 1970) has reported several important findings using the BT procedure: (*a*) *S*s consistently use hypotheses; (*b*) *S*s rarely change hypotheses after having received a positive OT; and (*c*) *S*s nearly always change hypotheses after receiving a negative OT. All of these results are consistent with the data reported by White (1972a, 1972b). First, Levine's *S*s typically receive extensive pretraining. Indeed, Levine assumes the universe of hypotheses is known by *S* and that the assumption can be met by training and instruction. White's (1971) finding was that *S*s receiving extensive systematic pretraining and instruction are likely to behave as hypothesis testers. Second, the behavior of extensively and systematically trained *S*s was as predicted by an error-trial-only CI model; *S* retains his hypothesis after positive outcomes and returns it to a pool of hypotheses for further sampling after negative outcomes. However, the question of whether the pretraining and instruction induced *S*s to develop and test hypotheses and thereby behave as they did in the BT problems has not been answered. It was, therefore, the purpose of the experiments reported herein to explore the effects of varying amounts of systematic training on performance in BT problems. Would, for example, highly trained *S*s provide more response sequences during BTs that are identifiable with hypotheses than untrained *S*s?

Experiment I

Method

Subjects and design. A total of 80 *S*s were randomly assigned, except for equating the number of males and females, to one of fiv[illegible]perimental conditions. The five conditions w[illegible] defined in terms of the number of pretraining [illegible]ms which preceded the administration of four [illegible] problems. The *S*s received either 0, 1, 2, 3, o[illegible] pretraining problems.

Stimuli. The stimuli consisted o[illegible] letters *X* and *T*, which were paired on each [illegible]lus card. The letters also varied in size (small [illegible]rge), texture (clear or crosshatched), and po[illegible]n (left or right side of card). The stimuli we[illegible]pared as 8 × 11 in. dittos.

A randomized blocking procedur[illegible]ed for arranging the pretraining stimuli. [illegible] consisted of the random arrangement of t[illegible] timuli of the population. Each dimension wa[illegible]evant to problem solution equally often. A ra[illegible]n order of the problems was selected for each [illegible]

Four of the aforementioned stimuli [illegible]lected for use during the BT problems. The [illegible]n was conducted such that certain configur[illegible] of responses during the BTs could be ind[illegible] consistent with one of eight unidimensiona[illegible]eses. The remaining four stimuli were used [illegible] the OTs. The OT stimuli were randoml[illegible]d for each problem with the restriction that e[illegible] appear equally often.

Procedure. The *S*s were run individua[illegible] by one of three *E*s. After being seated across a t[illegible]le from *E*, *S* was given routine instructions rega[illegible]ng the pretraining problems. Each *S* was shown [illegible] stimulus population, instructed as to the me[illegible]ng of feedback signals, and told that either text[illegible] position, size, or form of figure would defin[illegible]oblem solution.

For the various pretraining condition[illegible], an attribute for the single problem condition, or a combination of attributes for a series of problems, was randomly sampled. No repetitions of relevant attributes was permitted. The criterion of problem solution for these problems was set at 10 consecutive correct responses. The *S* was given a total of 48 trials to solve each problem, the problem being terminated by *E* showing *S* the solution. A trial consisted of showing *S* the stimulus card for approximately 5 sec., a pointing response by *S*, indicating which of the two figures contained the relevant attribute, and a verbal "right" or "wrong" by *E*. A period of approximately 5 sec. transpired between trials. The *S*s were asked to state the solution upon meeting criterion.

Following successful completion of pretraining, all *S*s were instructed as to the procedure involving BTs. They were told that feedback would not be given consistently as it had during pretraining, and were asked to continue responding as they would if feedback were given on each trial. Following these instructions, a series of four 15-trial problems was administered. For each of these problems, Trial 1 was designated as a positive OT. Regardless of *S*'s response, it was called correct. Trials 6 and 11 were also OTs. Feedback for the latter OTs was arranged such that one trial was designated positive

and the other negative. Since there are 16 possible combinations of feedback for Trials 6 and 11 for four problems, each *S* was randomly assigned to one of the sequences with the restriction that a single *S* be assigned to each order.

Results

According to Levine's (1966) hypothesis theory, a large percentage of the response sequences during BTs should be consistent with a single hypothesis. For his problems, approximately 91% of the sequences were found to be consistent with single hypotheses. In the present experiment, the proportion of univariate hypotheses was computed for each *S*, the means for the 0-, 1-, 2-, 3-, and 4-problem conditions (Conditions 0–4) being .69, .85, .86, .86, and .89, respectively. An analysis of variance applied to these data revealed significant differences among these proportions, $F\ (4, 75) = 3.86$, $p < .01$.

One major assumption made by certain hypothesis-testing theories of CI (e.g., Restle, 1962) is that *S* retains the hypothesis which dictated his previous response if that response is called correct. Indeed, Levine (1966) reported that hypothesis retention after correct responses occurred in 92% of the cases studied. Comparable proportions from the present experiment for zero through four problems, respectively, were .66, .72, .76, .81, and .75. The differences among these proportions were not reliable, $F\ (4, 75) = 1.24$, $p > .05$.

A second assumption made by the aforementioned hypothesis-testing theory, and by Levine (1970), is that *S* returns his hypothesis or set of hypotheses to a pool for resampling if the response dictated by the set is called wrong. Therefore, the probability of resampling that hypothesis should be low if *S* resamples from the available pool of hypotheses randomly, and exceedingly low if *S* remembers hypotheses that have been infirmed. Levine (1966) reported that on 98% of the occasions after *S*'s response had been called wrong a hypothesis different from the one manifested prior to the OT was observed. Again, for the present experiment, comparable values for Conditions 0 through 4 were .73, .91, .88, .86, and .94, respectively. Analysis of variance showed the differences among these proportions to be significant, $F\ (4, 75) = 2.54$, $p < .05$. Inspection of the proportions shows the difference to lie between Condition 0 and the other four conditions.

Since the frequency of hypothesis changes is a function of the number of pretraining problems administered, it might also be expected that similar observations could be made over the BT problems themselves. For that reason, the proportion of hypothesis changes after correct responses was computed for each BT problem, the obtained values for one through four BT problems, respectively, being .29, .36, .21, and .26. The obtained proportion of hypothesis changes after errors for one through four BT problems, respectively, were .85, .81, .91, and .90. Inspection of these da a indicate small differences and the lack of a discernible trend. Apparently, three OTs is an insufficient inducement for strategic change.

A natural extension of the assumptions listed above is that the size of *S*'s hypothesis pool remains constant regardless of the point at which the pool is measured. For example, if the pool is determined to contain, say, four hypotheses after the first negative OT, it ought to contain a like number of hypotheses after, say, the third consecutive negative outcome. To assess the size of the hypothesis pool, Levine (1966) computed the obtained proportion of successes for a group of *S*s. Reasoning that the obtained proportion of correct hypotheses would ordinarily be computed by dividing the number of logically correct hypotheses, as determined by feedback, by the number of hypotheses in the pool, Levine was able to obtain an index of the size of the pool by dividing the obtained proportion of hypotheses into the number of logically correct hypotheses. Using an identical procedure, the size of the hypothesis pool after each OT was estimated for each experimental condition and is shown in Table 1. Several trends are apparent from these data: (*a*) the size of the hypothesis pool is reduced over OTs; (*b*) *S*'s efficiency in handling information

TABLE 1

ESTIMATED SIZE OF THE HYPOTHESIS POOL AFTER EACH OUTCOME TRIAL IN EXPERIMENT I

Trial sequence	Number of pretraining problems				
	0	1	2	3	4
+, BTs	7.31 (5.41–7.69)	4.57 (4.30–5.13)	5.56 (4.88–6.78)	4.74 (4.40–5.56)	4.65 ([illegible]30–5.41)
+, BTs, +, BTs	4.93 (3.45–7.69)	2.37 (2.15–2.94)	2.75 (2.35–3.57)	2.75 (2.35–3.57)	2.52 ([illegible]–3.28)
+, BTs, −, BTs	6.40 (4.08–10.53)	3.37 (2.70–4.76)	3.65 (2.86–5.26)	3.26 (2.60–4.55)	3.88 ([illegible]–4.44)
+, BTs, +, BTs, −, BTs	4.00 (2.38–7.69)	2.91 (2.00–4.76)	2.20 (1.64–3.45)	1.65 (1.32–2.27)	[illegible]20 ([illegible]–3.45)
+, BTs, −, BTs, +, BTs	3.10 (2.00–5.56)	1.60 (1.70–2.22)	1.55 (1.27–2.13)	1.29 (1.14–1.67)	2.00 ([illegible]–2.94)

Note. Ninety-five percent confidence intervals appear in parentheses. Abbreviations: BTs = blank trials, + = positive outcome trial, − = negative outcome trial.

improves as a function of amount of training; and (*c*) the information from positive OTs is used more efficiently in reducing potentially valid hypotheses than that from negative OTs.

EXPERIMENT II

Experiment I showed reasonably strong training effects. In general, as *S* received more pretraining, he was more likely to alter his hypothesis after having received a negative OT and to maintain his hypothesis after positive OTs. Furthermore, *S* used both outcomes more efficiently with more pretraining in the sense that the size of the pool of hypotheses being entertained was smaller after, say, a single positive outcome for *S*s who had received three pretraining problems than for those who had received none. However, two aspects of the data were not entirely consistent with these aforementioned general trends and Levine's (1966) data. With four pretraining problems, *S*s were more likely to alter their hypotheses after positive OTs than *S*s with three pretraining problems. With respect to Levine's data, the probability of hypothesis change after positive OTs was considerably higher than for Experiment I. One viable explanation for the discrepancy is that different procedures were used. Levine, for example, began some of the BT problems with negative OTs, while all of the BT problems in Experiment I were begun with a positive OT.

Experiment II was designed to assess the reliability of the results of Experiment I and to determine what, if any, effects were produced by the variant procedures of Experiment I and Levine's (1966) experiment. The procedures of Experiment II were, therefore, as nearly identical to Levine's as was possible. In addition, a hypothesis arising out of Experiment I was tested by the manipulation of instructions. One interpretation of the high rate of hypothesis alteration after positive OTs is that *S* entertains several hypotheses simultaneously, and any one of these active hypotheses may be used as a basis for responding after a positive OT. To assess the accuracy of this interpretation, special instructions were included for some of the groups of Experiment II, asking them to consider a single hypothesis while attempting to solve the problem.

Method

Subjects and design. The *S*s were 128 male and female undergraduate students at Miami University who were assigned randomly, but in equal numbers with respect to sex, to one of eight experimental conditions. The conditions were defined by a factorial arrangement of two variables: number of

pretraining problems (0, 2, 4, or 6) and type of instructions (routine vs. restricting).

Materials and procedure. The materials and procedures were nearly identical to those of Experiment I save for the special instructions given to restrict *S*'s hypothesis pool and the designation of the various OTs as positive or negative.

The *Ss* in the special instruction condition were told to "consider only *one* characteristic as possibly relevant at any one time," and to "never, for example, consider the possibility that *X* and/or *large* are the relevant characteristics."

The dimensions made relevant to solution were balanced for frequency of occurrence and order to the extent that this was possible. For the two-problem condition, the 12 permutations of the four stimulus dimensions were assigned to the first 12 *Ss* assigned to that condition. The four dimensions were represented equally often in the problems assigned [illegible] other 4 *Ss* in the two-problem condition [illegible] the four-problem condition, four 4 × 4 Latin squares were selected to determine order of problem assignment. The Latin squares selected for the four-problem condition were also used in the six-problem condition. One of the permuted pairs or one of the remaining pairs of the two-problem condition was added to each row of each square to form the six-problem condition.

For the BT problems, eight sequences of positive (+) and negative (−) feedback were possible (e.g., + + +, + + −, etc.). Four 4 × 4 Latin squares were combined to determine the feedback sequences for each of the BT problems for the 16 *Ss* in each condition. The last OT of each BT was called positive to assure that *S* would not "leave" the univariate hypothesis realm.

Results

The mean proportion of hypothesis changes after negative OTs for each experimental condition are shown in Table 2. Analysis of variance applied to these data showed only a significant effect due to number of pretraining problems, F (1, 120) = 6.40, $p < .01$. The proportion of hypothesis changes was an increasing function of the number of pretraining problems.

Table 2 also displays the mean proportion of hypothesis changes after positive OTs. Analysis of variance showed the Type of Instructions × Number of Pretraining Problems interaction to be the only significant source of variation, F (3, 120) = 3.20, $p < .05$. Hypothesis changes were especially prevalent where *S* had not received instructions to limit his hypothesis pool to univariate hypotheses, especially in the case where two or fewer pretraining problems had been administered.

The mean proportions of multivariate hypotheses for each experimental condition are shown in Table 3. Any response sequence that could not be identified as consistent with a known univariate response sequence was assumed to be the product of a multivariate hypothesis. Analysis of variance applied to these data revealed significant effects due to type of instructions, F (1, 120) = 10.84, $p < .01$, with the groups being instructed to restrict their hypothesis pool to univariate hypotheses showing a smaller proportion of multivariate hypotheses than the noninstructed group. Additionally, moderately reliable effects were observed in the Instruction × Pretraining Problem interaction, F (3, 120) = 2.53, $p < .10$. As the number of pretraining problems increased, the proportion of multivariate hypotheses was reduced for the routine instruction condition. The proportion of multivariate

TABLE 2

Mean Proportion of Hypothesis Changes after Negative and Positive Outcome Trials in Experiment II

Instruction	Number of pretraining problems: 0	2	4	6
Negative outcome trials				
Routine	.75	.95	.94	.96
Restricted	.84	.90	.96	.96
Positive outcome trials				
Routine	.36	.48	.25	.21
Restricted	.31	.17	.19	.34

TABLE 3

Mean Proportion of Multivariate Hypotheses for Each Experimental Condition

Instruction	Number of pretraining problems: 0	2	4	6
Routine	.25	.15	.12	.14
Restricted	.08	.07	.09	.13

TABLE 4

ESTIMATED SIZE OF THE HYPOTHESIS POOL AFTER NEGATIVE OUTCOMES ON TRIALS 1, 2, AND 3 IN EXPERIMENT II

Number of problems	Outcome trial		
	1	2	3
0	6.90 (5.80–8.70)	6.25 (4.55–9.09)	4.55 (3.57–5.26)
2	6.35 (5.48–7.84)	4.55 (3.64–6.06)	4.35 (2.94–6.67)
4	6.06 (5.26–7.40)	3.57 (2.99–4.55)	3.03 (2.22–4.35)
6	5.80 (5.06–7.14)	3.45 (2.90–4.35)	2.44 (1.89–3.33)

Note. Ninety-five percent confidence intervals appear in parentheses.

hypotheses was nearly constant for the special instruction condition.

As in Experiment I, no systematic effects were observed as a function of BT problem.

An estimate of the size of the hypothesis pool after negative outcomes on OT_1, OT_2, and OT_3 as a function of the number of pretraining problems is shown in Table 4 for the routine and restricted instructional conditions, respectively. The estimates were arrived at by computing the proportion of correct hypotheses after each negative OT for each condition and dividing the appropriate proportion into the number of logically correct hypotheses (i.e., four after OT_1, two after OT_2, and one after OT_3) for the trial involved. In general, these data show the size of the hypothesis pool to be a decreasing function of the trial involved, that the size of the pool is reduced more substantially from trial to trial as *S* has more experience with the type of problem at hand, and that the size of the pool is smaller in the case where *S*s were specially instructed to restrict their hypotheses to a univariate realm.

TABLE 5

ESTIMATED SIZE OF HYPOTHESIS POOL AS A FUNCTION OF NUMBER OF PREVIOUS POSITIVE OUTCOME TRIALS (OTs) AFTER NEGATIVE OT_3

Number of problems	Number of preceding positive OTs		
	0	1	2
0–2	6.25 (3.13–12.50)	4.00 (2.63–6.67)	2.50 (1.92–5.00)
4–6	2.63 (1.82–4.35)	3.57 (2.50–5.56)	2.13 (1.59–3.23)

Note. Ninety-five percent confidence intervals appear in parentheses.

An estimate of the size of the hypothesis pool as a function of the number of preceding positive OTs is shown in Table 5. The estimates were obtained by computing the proportion of correct hypotheses after a negative outcome on OT_3 and dividing the obtained proportion into the number of logically correct hypotheses after OT_3 (i.e., one). Unfortunately, there were no discernible trends where the data were analyzed separately for each condition. Quite likely, too few data points (e.g., in most cases, as few as eight) were involved in the estimates. To improve the reliability of the estimates, the data for the one- and two-pretraining-problem conditions (collapsed over instructions) were combined and compared with the combined results of the four- and six-problem conditions. In general, Table 5 shows a systematic reduction in the size of the hypothesis pool as a function of the number of preceding positive OTs for the zero–two problem combination, but no reduction in size of the pool for the four–six problem combination.

DISCUSSION

Hypothesis theory. Levine (1966, 1970) reported that *S*s attempting to solve simultaneous discrimination problems for which solution is defined by a single cue employ univariate hypotheses, usually alter their hypotheses after negative OTs, and rarely alter their hypotheses after positive OTs. An additional assumption is that *S* samples a subset of the total pool of potentially correct hypotheses. Both experiments appear to justify the first three assumptions, but only under certain conditions. The *S*s who receive relatively extensive pretraining on a problem identical to the test problem nearly always alter their hypotheses after negative outcomes, rarely alter them after positive outcomes, and restrict their hypotheses to the univariate realm. The *S* who receives very little pretraining either entertains multivariate hypotheses, or no hypotheses at

all, and is [illegible] to alter his hypothesis after both negative and positive outcomes.

The subset-sampling assumption appears to be correct in view of the present data. The probability that S would alter his hypothesis after a positive OT was much higher when instructions did not require him to restrict the size of his working hypothesis set to one. Moreover, the effect of such an instructional set is uniform over varying numbers of pretraining problems.

Finally, Levine (1966) reported that Ss used the information after positive OTs to adjust their hypothesis pools more than after negative OTs, but that Ss regularly reduced the working set of hypotheses after all outcomes. The results of the present set of experiments generally confirm Levine's observations, with one potentially important exception. In Experiment II, it may be worth noting that the size of the working hypothesis set is unrelated to the number of preceding positive OTs where Ss had received four or six pretraining problems. Unfortunately, the reliability of this finding is in question, since the results of Experiment I indicate that Ss do reduce the size of their working hypothesis set over positive OTs regardless of whether 0, 1, 2, 3, or 4 pretraining problems are given. Perhaps six or more pretraining problems would produce this effect with greater reliability. On an intuitive level, it seems plausible to assume that S would eventually adopt a strategy which produces a single hypothesis to be tested and retained, if positive OTs ensue, since S is not required to hold in memory a fairly large set of hypotheses, some of which have been confirmed and others that have not been tested. In any case, it is not clear at this point whether S adopts a strategy which places less emphasis on memory as he solves more problems of a particular class.

Multiprocess theory. The main assumption of the multiprocess point of view is that S is capable of several, perhaps many, approaches to a problem. Furthermore, it is assumed that, among other factors, the amount of experience S has had with the problem at hand determines the approach he will use in solving the problem. The results of both experiments appear to be entirely consistent with these general assumptions. In addition, the notion expressed by White (1972a, 1972b) that inexperienced problem solvers would adopt an associative learning strategy, while more sophisticated Ss would adopt a more efficient and elegant strategy of testing hypotheses, appears to have been confirmed by the present data.

More specific with respect to the associative-to-hypothesis learning progression are the data of untrained Ss who show a high rate of unidentifiable BT sequences, a potential indicator of nonhypothesis behavior on their part. However, with training the rate of such BT sequences was reduced, but still remained above 10%, indicating that Ss were not simply dropping multidimensional hypotheses from their working pool.

A corollary of the notion that S will adopt a hypothesis-testing strategy with training is that rarely will S alter his hypothesis after a positive OT. In fact, the rate of hypothesis change was reduced over training problems; in addition, the size of the hypothesis set remained relatively constant over a series of positive OTs for extensively trained Ss, a result which is consistent with the one-hypothesis-at-a-time model of Restle (1962).

A final major hypothesis predicted that the rate of hypothesis alteration after negative OTs would be an increasing function of number of pretraining problems, and that the probability of hypothesis change after positive and negative OTs would be equal. The first half of the hypothesis was confirmed by the data; hypothesis change occurred on 96% of the trials for the six-problem condition of Experiment II and was somewhat lower for all of the remaining conditions of both experiments. However, the second half of the hypothesis was not supported by the data in any condition. One interpretation of the fact that the rate of hypothesis change was considerably higher after negative OTs than positive OTs has been suggested by White (1972a). In the latter paper, it was hypothesized that within a sample of adult human problem solvers a variance of strategies would obtain. For example, from instruction and previous problem-solving experience, some Ss would decide that their task was to memorize the response that was to be given in the presence of each stimulus; no consideration would be given to the possibility that the various stimuli are related. Other Ss might conceptualize the task as requiring an active reorganization of the problem, and the formulation and testing of hypotheses. A proper mix of such problem solvers could conceivably have produced the data obtained in both experiments. If the data of a sizable number of Ss who are testing hypotheses, and who are altering their hypotheses after errors and maintaining con-

firmed hypotheses are combined with a group of *S*s who are adopting an associative strategy, and who are altering "hypotheses" after both positive and negative OTs, it would be expected that the rate of hypothesis alteration would not be equal after positive and negative OTs, and that some moderate rate of hypothesis change would occur after positive and negative OTs. While the notion of variant strategies among *S*s has not been directly determined by the present experiments or White's (1972a, 1972b) earlier experiments, the hypothesis appears more credible in light of these experiments.

Theoretical sketch. The data from the present and earlier experiments suggest that *S* has available several approaches to a CI problem. The most rudimentary approach simply involves the representation of stimuli in their entirety with a proper response. Given the known difficulty *S*s have with negative instances of concepts, it seems plausible to assume that *S* organizes the instances in such a way as to differentiate the negative and positive instances; however, the negative instances are ignored so far as solution to the problem is concerned. If *S* finds that simple representation of the stimuli and responses is overloading memory, it is likely that he will adopt a new strategy. If such an interpretation is correct, it would be expected, for example, that training effects would be far less substantial if the stimulus population was delimited to a few examples. Moreover, a nonassociative strategy would be more likely to occur if *S* knew the stimulus population to be large (e.g., selection paradigm).

Movement to the nonassociative or hypothesis-testing realm, then, depends on the knowledge *S* has of the problem at hand and the development of the hypothesis-testing skill itself. The data that are available seem to indicate that Levine's (1966) description of the hypothesis-testing *S* is correct at least in those cases where *S* knows it is clearly beneficial to him to consider several hypotheses simultaneously and to retain in memory those that have been infirmed. It could also be assumed that the hypotheses generated early in a series of problems are restricted to those suggested by positive examples. Later in training, negative examples may serve as a source for hypothesis generation and hypothesis testing. Finally, *S* may adopt the single-hypothesis selection mode if circumstances are appropriate. For example, in the present experiments, *S* learns that 16 trials are given for each problem, with four outcomes per problem. Since [illegible]iency is not being reinforced, movement to[ward] a less taxing strategy would be expected.

So far as training effects, [illegible] changes, and the development of the ability to test hypotheses are concerned, it seems the most powerful test of the overall hypothesis would be with young children (up to [illegible] yr.), who are known not to possess the ability to test hypotheses (e.g., Douglass & Bourne, 1971). Training effects would not be expected for these *S*s.

REFERENCES

Bourne, L. E., Jr., & Restle, F. A mathematical theory of concept identification. *Psychological Review*, 1959, **66**, 278–296.

Bower, G. H., & Trabasso, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford, Calif.: Stanford University Press, 1964.

Douglass, H. J., & Bourne, L. E., Jr. Chronological age and performance on problems with repeated presolution shifts. *Developmental Psychology*, 1971, **4**, 329–333.

Levine, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, **71**, 331–[illegible]

Levine, M. Human discrimination learning: The subset-sampling assumption. *Psychological Bulletin*, 1970, **74**, 397–404.

Restle, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 320–343.

White, R. M., Jr., Richards, D. L., & Reynolds, R. Number of pretraining problems and type of instructions in concept identification. *Psychological Reports*, 1971, **28**, 643–649.

White, R. M., Jr. Effects of some pretraining variables in concept identification. *Journal of Experimental Psychology*, 1972, **94**, 198–205. (a)

White, R. M., Jr. Relationship of performance in concept identification problems to type of pretraining problem and response-contingent postfeedback intervals. *Journal of Experimental Psychology*, 1972, **94**, 132–140. (b)

(Received October 3, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1061–1068

SEMANTIC ENCODING AND RECOGNITION MEMORY: A TEST OF ENCODING VARIABILITY THEORY[1]

EUGENE WINOGRAD[2] AND MARY FULCHER GEIS

Emory University

periment I compared recognition memory for two types of homographs. balanced homographs, there are two relatively equiprobable semantic odings; for polarized homographs, there is one dominant encoding. It was edicted, on the basis of encoding variability theory, that recognition would e superior for polarized homographs but the opposite outcome was obtained it two retention intervals and replicated. In Experiment II, successive word ssociation tests were administered at two intertest intervals. The semantically ored responses supported the assumption of greater variability of semantic codings for the balanced homographs. A retrieval strategy of semantic oding was proposed to account for the better recognition of balanced mographs.

The present research is an extension of encoding variability theory to the semantic encoding of ambiguous words. According to Martin (1968) and others, the subject makes a perceptual-encoding response to a presented stimulus. Although this encoded, functional stimulus may not be identical to the nominal stimulus, it constitutes the memorial representation of the nominal stimulus. The stimulus-as-coded (Bower, 1972), not the stimulus-as-presented, is stored in memory. Encoding variability theory further assumes that a given *S* may make different encodings of the same nominal stimulus on different occasions. Since Martin's research has been concerned with paired-associate learning and transfer situations in which trigrams were used as stimuli, he has viewed encoding as stimulus selection. The principles of encoding variability theory are, however, applicable to a wide variety of experimental contexts.

The experimental context with which we are concerned is recognition memory for familiar English words, and the type of encoding with which we are concerned is the memorial representation of the semantic features of a stimulus. One basis for the presumed encoding variability of familiar words is their semantic ambiguity. In a large number of instances, the same sequence of letters and phonemes may have two or more quite different meanings. Examples of such polysemous words, or homographs, are BANK, which may refer to a financial institution as well as to land adjoining a body of water, and CLUB, which may refer to an organization of people united by a common interest and to a wooden stick used as a weapon. The distinct meanings of a homograph may be considered different semantic senses of the words that allow for different semantic encodings of the same nominal stimulus across *S*s in a psychological experiment, or more importantly here, by the same *S*s on different occasions.

We believe that encoding variability theory has clear implications for recognition memory research. To recognize an old item correctly, it is assumed that *S* must make the same encoding of the item at the time of study and at the time of test. If the test encoding of an item differs from the study encoding of the item, *S* will not recognize the item, since he is accessing an encoding that does not have appropriate occurrence information, whatever that may be, associated with it. From this perspective, forgetting may be explained as failure of trace contact due to differences between

[1] This research was supported by National Science Foundation Grant GB-18703. The second author was supported by a National Defense Education Act Title IV fellowship from the Office of Education, U.S. Department of Health, Education and Welfare. The assistance of Mary Pender and Deva Scheel is gratefully acknowledged.

[2] Requests for reprints should be sent to Eugene Winograd, Department of Psychology, Emory University, Atlanta, Georgia 30322.

the stored encoding and the test encoding, rather than as loss of stored information.

Another aspect of the present argument which needs to be made explicit is an assumption of encoding specificity. According to this assumption, only a single functional encoding is made by the subject on a given occasion, with an occasion defined as the presentation of a verbal stimulus on either a study or a test trial. Light and Carter-Sobell (1970) and Winograd and Conn (1971) have presented evidence consistent with the specificity assumption as it applies to the semantic encoding of homographs.

The first experiment and its replication are concerned with recognition memory for two types of homographs. A *balanced* homograph is defined here as having two distinct semantic senses that are relatively equiprobable encodings. A *polarized* homograph is defined as one with two distinct semantic senses that are markedly different in their semantic encoding probabilities, with one sense being highly likely. In our opinon, encoding variability theory leads to the prediction that recognition memory for polarized homographs should be superior to recognition memory for balanced homographs. With high likelihood, *S* should make the same semantic encoding of a polarized homograph at study and at test, since its one dominant sense is a highly probable encoding. As argued above, trace contact and successful recognition depend upon such identity of study and test encodings. On the other hand, since two senses of a balanced homograph are equally probable semantic encodings, *S*'s encoding of a balanced homograph may vary from one occasion to the next. When different encodings are made at study and at test, the consequence is failure to recognize the item as old. Thus, successful recognition of old items should be more likely for polarized than for balanced homographs.

The longer the interval between study and test, i.e., the longer the interval between successive encodings of an item, the greater should be the likelihood that the encodings are independent. If this hypothesis of increased encoding variability over time is correct, the predi[illegible]riority of recognition memory for [illegible]ed homographs should be most ev[illegible]t a long retention interval.

Experiment [illegible]

Method

Design. The design of Experim[illegible] Ia and Ib was a 2 × 2 factorial in which type [illegible] homograph, balanced or polarized, was a wit[illegible]jects variable and retention interval was a [illegible]n-subjects variable. The length of the ret[illegible]tervals is the only difference between the [illegible]periments. In Experiment Ia, the retention [illegible]ls were 3 min. and 25 min.; in Experiment I[illegible] retention intervals were 3 min. and 48 hr.

Materials. The homographs and [illegible] semantic encoding probabilities were taken [illegible]e norms of Cramer (1970), Kausler and Kolla[illegible]970), and Perfetti, Lindsey, and Garson (1971[illegible] majority of the homographs were selected f[illegible] the last-named source. These norms had be[illegible]ined by requiring *S*s to give a single word ass[illegible]tion to a series of homographs. Judges classifi[illegible] associations according to the semantic se[illegible]at they represented. For example, associat[illegible] such as "money," "vault," and "check" to the homograph BANK were classified as representing the encoding of bank as a financial institution; associations such as "river," "hill," and "slope" were classified as representing the encoding of "land adjoining a body of water."

Fifty-six balanced and 56 polarized homographs[3] were used in Experiment I. The selection criterion for a balanced homograph was that fewer than 75% of the associations were classified by the judges under the primary semantic sense (S_1). In other words, the normative or cultural encoding probabilities of S_1 for the balanced homographs were less than .75. For the 56 balanced homographs, the mean probabilities of S_1 and S_2, the secondary semantic sense, were .54 and .31. The selection criterion for a polarized homograph was that at least 75% of the associations were classified by the judges under the primary semantic sense The mean probabilities of S_1 and S_2 for the polarized homographs were .83 and .11. The median word frequencies (Kučera & Francis, 1967) were 25 per million for the balanced homographs and 36 per million for the polarized homographs. The corresponding means were 99 and 107 per million.

Two lists, A and B, each containing 28 balanced and 28 polarized homographs, were constructed. One random order of each list was used for study. The list that was not presented for study supplied the distractor items on the two-alternative, forced-choice recognition test. Three different test forms were constructed by randomly pairing List A and List B homographs that were of the same type, i.e.,

[3] A list of the words used in Experiment I is available from the first author.

on the test, the choice was always between two balanced homographs or two polarized homographs. Since list effects were not significant, this variable is not mentioned again.

Procedure. Subjects were shown either the List A or List B slides at a 3-sec. rate. In both experiments, approximately half of the *S*s who were tested at each retention interval studied List A. Following presentation of the last word, *S*s worked for 3 min. on number progression problems. In Experiment Ia, half of the *S*s were administered the recognition test immediately after the filler task. Before the other *S*s were administered the recognition test, they received a 22-min. break during which they were asked to leave the room and not to discuss the experiment. In Experiment Ib, the filler task was followed by either immediate testing, as in Experiment Ia, or testing at 48 hr. The experiments were conducted in sessions with small groups of *S*s. Experiment Ia was completed before Experiment Ib was begun.

Subjects. Fifty *S*s participated in the first experiment and [illegible] *S*s participated in the second experiment. [illegible] were assigned randomly to lists, test forms, and retention intervals, until there were 25 *S*s at each interval in Experiment Ia and 24 *S*s at each interval in Experiment Ib. Subjects in Experiment Ia were summer session students at Emory University. [illegible] 41 were paid for participation and 11 were fulfilling an introductory psychology course requirement. Subjects in Experiment Ib served during the regular school year to meet an introductory psychology course requirement.

Results and Discussion

The results of the experiments are shown in Table 1. In Experiment Ia, the main effect of type of homograph was the only significant outcome, $F\ (1, 48) = 21.60$, $p < .001$.[4] The interesting aspect of the results is that recognition memory for balanced homographs was superior to recognition memory for polarized homographs. Since this result was contrary to predictions, and since no effect of retention interval was evident, the experiment was repeated with a retention interval that was sufficiently long to allow for a performance decrement.

In Experiment Ib, recognition memory for balanced homographs was again superior to that for polarized homographs, $F\ (1, 46) = 20.41$, $p < .001$. There was a significant decline in performance at the 48-hr. retention interval, $F\ (1, 46) = 18.55$, $p < .001$, but there was no statistical support for an interaction between type of homograph and retention interval.

TABLE 1

CORRECT RECOGNITION PROPORTIONS

Experiment	Type of homograph: Polarized	Balanced
Ia		
3 min.	.80	.86
25 min.	.83	.87
Ib		
3 min.	.81	.86
48 hr.	.65	.76

Imagery ratings of the homographs were collected to determine whether an uncontrolled difference in rated imagery might account for the difference in recognition memory between the two types of homographs. Thirty-seven undergraduates from the Emory University subject pool rated each homograph on a 7-point imagery scale, according to the procedure described by Paivio, Yuille, and Madigan (1968). The difference between the mean rating of the balanced homographs (4.94) and of the polarized homographs (4.90) was not significant.

The data of Experiments Ia and Ib evidence a relationship between semantic encoding probabilities and performance in a recognition memory task. Since the direction of the relationship was contrary to predictions, it is necessary to reconsider our argument. Although this reconsideration will be undertaken more fully after presentation of Experiment II, one hypothesis that led to the second experiment is mentioned here. It is conceivable that *S*s are more likely to encode both semantic senses when there are two encodings of similar strength, as is the case with balanced homographs, than when there is one culturally dominant encoding, as with polarized homographs. If two semantic representations are encoded, then either encoding at the time of test can result in trace contact and successful recognition. While such double encoding violates the strong form of

[4] All *F* values reported in this paper treat words as a fixed effect. If words are treated as a random factor and *min F′* is computed, following Clark (1973), all of the significant outcomes remain so.

the encoding specificity assumption outlined earlier, it is consistent with a weak form of encoding specificity, namely, that one or more semantic representations may be tagged on a given occasion. As Light and Carter-Sobell (1970) point out, this assumption is still distinguishable from an assumption of generic encoding, since it is not assumed that all senses sharing the same lexical form are encoded whenever that form is studied.

Mnemonic performance in Experiment I was predicted from encoding data obtained by other investigators. These encoding data were taken from word associations given on a single occasion, while our concern is with variability within *S*s on two occasions. Therefore, Experiment II was conducted to assess the stability of semantic encodings over the time periods used in the first experiment with the expectation that such data would illuminate the recognition memory performance of Experiment I.

Experiment II

Method

Design. Two coordinate procedures were employed. In one procedure, discrete word associations to the same homographs were obtained from the same *S*s on two occasions separated by either 5 min. or 48 hr. In the other procedure, the method of continued association was used to obtain from *S*s two different associations to the same homographs on one occasion. From the first procedure, data concerning the relative stability of semantic encodings over time were obtained. From the second procedure, the likelihood of *S* making two semantically distinct encodings on the same occasion, when instructed to produce two different responses, was determined.

Materials. More stringent selection criteria were used to choose 25 balanced and 25 polarized homographs from the norms cited in Experiment I. A homograph was considered balanced if the normative encoding probability of S_1 was less than .58 and the normative encoding probability of S_2 was at least .40. The mean published values of S_1 and S_2 for the balanced homographs were .49 and .41. A homograph was considered polarized if the normative encoding probability of S_1 was at least .78 and if the normative encoding probability of S_2 was at least .10. The mean published values of S_1 and S_2 for the polarized homographs were .81 and .14. The actual S_1 and S_2 values for the Emory population, as distinguished from the published values, are reported in the results section. Thirty nonhomographs were also selected, half of which were used in the booklets for the first [illegible] and the other half of which were used in th[illegible]klets for the second test at 5 min. or 48 hr. T[illegible]homographs were chosen from words whose di[illegible]ary (Random House Dictionary of the English [illegible]guage, 1968) meanings represented only one se[illegible] sense in the authors' judgment. The m[illegible] frequencies (Kučera & Francis, 1967) were 28 [illegible]illion for the balanced homographs, 30 per mill[illegible]or the polarized homographs, and 29 per [illegible]on for the nonhomographs.

In the booklets for the discr[illegible]iation procedure, a single word appeared on [illegible]page with a space for *S*'s response next to th[illegible]ls. In the booklets for the continued associ[illegible] procedure, the same word appeared twice on [illegible]ge with a space for *S*'s response next to each o[illegible]rence of the word. Each *S* received a 67-page bo[illegible]. A space for *S*'s name and instructions were [illegible]ed on the first 2 pages. The remaining 65 pa[illegible]onsisted of the 25 balanced homographs, the 25 [illegible]zed homographs, and 15 of the nonhomograph[illegible]ers. Four random orders of the words were used [illegible]struction of the booklets

Procedure. Small groups of *S*s were t[illegible]ed. Subjects in the discrete association proc[illegible] were instructed to write down the first word [illegible]t came to mind upon seeing a stimulus word. A[illegible] completing the first association test, all *S*s in [illegible]e discrete association procedure received a 5-[illegible] number progression task. In the 5-min. con[illegible], the *S*s then were readministered the associatio[illegible]st, while in the 48-hr. condition, the *S*s were d[illegible]ssed and retested 48 hr. later. For the second [illegible]sociation test, *S*s received a booklet consisting of t[illegible]50 homographs on which they had been teste[illegible]eviously and the 15 nonhomographs on which the[illegible] had not been tested previously. The words in the second test booklet were in a different random order from those in the first test booklet. The order of the words and the filler words were changed to reduce *S*'s tendency to respond exactly as before, i.e., to minimize possible intralist context effects. Subjects in the continued association procedure were dismissed upon completion of their one test booklet.

Subjects. The *S*s were 105 Emory undergraduate students who participated to fulfill an introductory psychology course requirement. Groups were randomly assigned to conditions and booklets. In the discrete association procedure there were 38 *S*s in the 5-min. condition and 30 *S*s in the 48-hr. condition. In the continued association procedure, 37 *S*s were tested.

Scoring. Two judges independently classified the associations according to the semantic senses that were represented. The semantic senses that were used in scoring were similar to those that were reported in the norms cited in Experiment I. There was 95% agreement between the two judges. A third judge resolved the scoring disagreements. The third judge and one of the first two judges were unaware of the purpose of the scoring and of the expected distributions of encoding probabilities.

Six percent of the associations were unscorable

in that [illegible]ges could not classify them under a reason[illegible]antic sense. These associations were exclud[illegible]he reported analyses. Three percent of the [illegible]g associations were scored as representative [illegible]hird semantic sense, i.e., 97% of the scorable [illegible]tions were classified under either S_1 or S[illegible]nd conditional probabilities for the tertiary [illegible]re not reported because of the small number [illegible]tions that were classified under S_3.

Results a[illegible]scussion

Stability [illegible]f group norms. The stability of the obtai[illegible]ed group norms as well as their comparabi[illegible]y with the norms of other investigat[illegible]s will be reported first. Based on the d[illegible] from the first test for the 68 *S*s in th[illegible]discrete association procedure, the ob[illegible] mean values of S_1 and S_2 for the 25 [illegible]rized homographs were .87 and .12. [illegible] th[illegible]5-min. retest (n = 38), the mean [illegible]f S_1 and S_2 were .88 and .12; at 48[illegible]30), they were .86 and .12. Thes[illegible] in good agreement with the [illegible]es of .81 and .14 that were obtai[illegible]n the published norms cited in E[illegible]nt I and also demonstrate the stabilit[illegible]ne group norms.

For 10 [illegible] the 25 balanced homographs, but for no[illegible] of the polarized homographs, there were discrepancies between the Em[illegible]y norms and the published norms as to which semantic sense was dominant, i.e., as to the definitions of S_1 and S_2. This outcom[illegible] not surprising, since the criterion for a balanced homograph was that its semantic encodings were relatively equiprobable. When S_1 and S_2 are redefined according to the obtained norms, the balanced homographs became more polarized than when they are based on the published values. At the first, 5-min., and 48-hr. tests, the mean S_1 and S_2 values were .63 and .34, .62 and .34, and .62 and .3[illegible] respectively. In subsequent analyses, the de[illegible]ns of S_1 and S_2 and their encoding probabilities are based on the Emory data.

Although the balanced and polarized homographs are different with respect to the probabilities of their primary associates when the associates are defined semantically, it should be noted that they are not different when the primary associate is scored in terms of words. In word association norms, the primary associate usually is defined as the single word that the subjects gave most frequently. For the present materials, collapsed across procedures and tests, the mean probabilities of the primary word associates were .26 for the polarized homographs and .24 for the balanced homographs.

TABLE 2

Encoding Patterns from Test to Retest

	Time	Polarized	Balanced
P(same)	5 min.	.88 (902)	.79 (886)
	48 hr.	.87 (710)	.77 (699)
$P(S_2 \vert S_1)$	5 min.	.06 (793)	.15 (552)
	48 hr.	.07 (616)	.17 (448)
$P(S_1 \vert S_2)$	5 min.	.44 (106)	.28 (310)
	48 hr.	.48 (86)	.28 (238)

Within-subject encoding stability. Data relating to the question of the stability of semantic encodings over time are presented in Table 2. The entries in Table 2 are for pooled observations over *S*s and words. The number of observations in each cell is given in parentheses. In the upper portion of the table, the probabilities of making the same semantic encoding at test and retest are shown. At each interval, the probability of the same semantic encoding is less for balanced homographs than for polarized homographs. The data were subjected to an analysis of variance by taking the proportions across words for each *S*. The effect of type of homograph was the only significant outcome, F (1, 58) = 52.38, $p < .001$. The assumption of greater variability of semantic encoding for balanced homographs, on which the first experiment was based, is supported. However, the assumption of increased variability over time is not supported.

The likelihoods of switching from S_1 to S_2, and vice versa, are shown in the middle and lower portions of Table 2. In the middle portion of the table, $P(S_2|S_1)$ is the probability of a switch to the secondary encoding at retest, conditionalized on the occurrence of the primary semantic encoding at the first test. Although the probabilities of switching from S_1 to S_2 are low, the greater likelihood of switching for the

balanced homographs is significant, $F(1, 58) = 44.61$, $p < .001$.

The bottom portion of Table 2 contains the probabilities of switching from S_2 to S_1, $P(S_1|S_2)$. These probabilities are substantially higher than those of the preceding analysis and are an indication that subjects are less likely to repeat an encoding if it is the normative secondary than if it is the normative primary. It is evident that the probability of an encoding switch from the normatively weaker to the normatively stronger encoding is greater for polarized homographs, $F(1, 52) = 10.39$, $p < .01$.

Independence of encodings. The question of whether responding at the second test was independent of responding at the first test was examined next. The observed joint probability of particular test–retest encoding patterns was compared to a predicted joint probability that was determined from independent estimates of S_1 and S_2 encoding probabilities. Since all values for 5 min. and 48 hr. were almost equal, these data were combined. For example, the observed joint probability of an S_1 encoding at both test and retest, $P(S_1 \text{ and } S_1)$, was .52 for balanced homographs. The predicted $P(S_1 \text{ and } S_1)$, if it is based on independent estimates of $P(S_1)$ on each occasion, is $P(S_1)P(S_1)$, or (.63) (.62), or .39. In every comparison, there was a lack of agreement between the predicted and observed values and the discrepancies were consistent. The observed probability of making the same encodings over time was greater than expected on the independence assumption, and the observed probability of two different encodings over time was smaller than expected. The differences are significant by chi-square. There is no evidence that semantic encoding on the second test is independent of the first test.

Scoring by word. So far, the data have been considered largely in terms of semantic types or senses. The probability of obtaining the identical word as the response to the same homograph at both tests is also of interest. For the polarized homographs, the probabilities of the same word at both tests were .54 at both 5 min. and 48 hr. For the balanced homographs, the corresponding values were .49 and .45. The effects of type of homograph, time, and their interaction were not significant. Semantic scoring apparently was not affected by differential bias to repeat particular words for either balanced or polarized homographs. This outcome was expected, since the strength of the primary *word* associate, as reported above, was the same for the two types of homographs.

The absence of a significant effect of test–retest interval on word repetition is consistent with the outcomes of semantic scoring. Our data are in close agreement with those of Fox (1970) with respect to within-subject change. Fox administered identical word association tests to the same *S*s with a test–retest interval of 59 days and scored the responses by exact words rather than by semantic senses. He reported that the probability of within-subject word change was .52. For our data, the overall probability of a word change was .48. Thus, across the two studies, the probability of *S* changing his associative response does not increase from 5 min. to 59 days. Both here and in Fox's work, there is evidence for considerable normative or within-group stability in conjunction with considerable within-subject instability. That is, while *S*s change their responses, the group norms remain constant.

Continued association. In the continued association procedure, the subjects were required to produce two successive but different associations to each homograph. The probabilities that the two associations represented the same semantic sense were .65 for the polarized homographs and .52 for the balanced homographs. The difference was significant, $F(1, 72) = 39.06$, $p < .001$. When the probability of a switch to the secondary encoding is conditionalized on the occurrence of the primary encoding as the first response, i.e., $P(S_2|S_1)$, the resulting values are .26 for the polarized homographs and .36 for the balanced homographs, $F(1, 72) = 19.27$, $p < .001$. For the opposite pattern, i.e., $P(S_1|S_2)$, the conditional probabilities are .75 and .54 for the polarized and balanced

homog...hs, respectively, F (1, 70) = 34.46, p < .001. The patterns are the same as those obtained for discrete associations (Table 2), but the likelihood of change ...ater in all cases for continued associ... This difference may be due to the ...traint of not being allowed to repeat the same word in the continued associat... procedure.

General Discussion

Experiment I was based on some assumptions about semantic encoding processes leading to the prediction that recognition memory for balanced homographs would be inferior to recognition memory for polarized homographs. However, the opposite state of affairs prevailed: recognition memory for balanced homographs was superior to that for polarized homographs. This finding was replicated, and the assumptions concerning the variability of semantic encoding of the two types of homographs ...ted by administering a word association test on two occasions. The word association findings were clear-cut: In agreement with the assumptions underlying the recognition memory experiments, semantic encoding was more variable for balanced than for polarized homographs.

Since the results of Experiment II support the assumptions on which the experimental tests of encoding variability theory were based, we must reconsider what is happening in recognition memory. Earlier, a model that incorporated the assumption of multiple encoding was mentioned as a possible fit for the recognition memory data. In a simple form of such a model, it is assumed that S may achieve, on some occasions, more than a single semantic encoding of a homograph. If multiple representations are encoded at study, recognition memory should be improved, since trace contact can result if any one of the previously stored encodings is made at test. The encoding data from the continued association procedure of Experiment II are evidence that, when more than one encoding is attempted, S is more likely to arrive at two semantically distinct encodings for balanced homographs than for polarized homographs. When Ss were required to provide two responses, the second response represented a different semantic sense significantly more often for the balanced homographs. Thus, the superior recognition of balanced homographs is consistent with an assumption of multiple encoding.

Rather than assume multiple semantic encoding at study, we consider it more likely that Ss in a word-memory experiment make a single encoding of an item at study but, when forced to choose between two seemingly unfamiliar items, attempt to recode the test items semantically at the time of test. (It is assumed that, if S makes a test encoding of an old word that differs from its study encoding, he is placed in the situation of having to choose between two unfamiliar items on a forced-choice recognition test.) Furthermore, Ss may be more likely to recode if they notice that the words in an experiment are homographs.

We do not know of any evidence to compel an assumption of recoding at test over the alternative assumption of multiple encoding (double tagging) at study. However, at present, we believe that it is both parsimonious and plausible to continue to assume encoding specificity at study, unless an alternative strategy is made more efficient by special conditions such as producing puns. This view preserves the original encoding variability argument on which this research was based, with the addition of an assumption concerning the likelihood of additional retrieval operations at test.

An alternative account of the superior recognition performance for the balanced homographs would be in terms of word frequency. Although the balanced and polarized homographs were equated on word frequency, it is obvious that they were equated only on frequency of the homograph form, rather than on frequency of particular meanings. For example, assume that the homographs BANK and CLUB are of equal frequency. If CLUB is balanced and BANK is polarized, then the primary encoding of CLUB necessarily will be lower in frequency than the primary encoding of BANK, since the primary sense of a balanced homograph must have a smaller slice of the frequency pie than the primary sense of a polarized homograph if the pies have been kept the same size. Thus, it is possible that the superior recognition of balanced homographs is a special case of the well-documented finding that recognition memory is better for low-frequency than for high-frequency words (Schulman, 1967; Shepard, 1967). Whether differences in frequency of the semantic types used here are sufficient to account for the difference found in recognition memory is an empirical question. The reader will note that the

selection of items matched in S_1 frequency but differing in S_1 and S_2 probabilities not only is a formidable task in itself, but necessarily must result in unequal frequencies of the homograph forms.

Finally we note some interesting features of the word association data of Experiment II. With regard to stability, a distinction must be made between scoring for repetition of the same word, the traditional word association measure, and scoring for repetition of the same category of semantic meaning. With respect to the exact associate, *S* is as likely to change his response as to repeat it. If our data are compared to Fox's (1970), this within-*S* instability is relatively constant from 5 min. to 59 days. If word associations are scored semantically, they are much more stable on the whole (see Table 2). But, even with semantic scoring, time between successive tests has very little effect, at least over a 48-hr. interval.

We seem to have assessed a relatively permanent characteristic of the semantic memory system. In distinguishing between semantic and episodic memory, Tulving (1972) points out that information in semantic memory is not temporally dated while information in episodic memory is. In these terms, the present research was an attempt to predict some simple phenomena of episodic memory, using a recognition memory task, on the basis of independently determined characteristics of semantic memory. Although our particular predictions were not confirmed, we have found a relationship between the two systems that must be accounted for.

REFERENCES

Bower, G. H. Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, D.C.: V. H. Winston, 1972.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–359.

Cramer, P. A study of homographs. In L. Postman & G. Keppel (Eds.), *Norms of word association.* New York: Academic Press, 1970.

Fox, P. W. Patterns of stability and change in behaviors of free association. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 30–36.

Kausler, D. H., & Kollasch, S. F. Word associations to homographs. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 444–

Kučera, H., & Francis, W. N. *Computational analysis of present-day American English.* Providence: Brown University Press, 1967.

Light, L. L., & Carter-Sobell, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 1–11.

Martin, E. Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review*, 1968, **75**, 421–441.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76** (1, Pt. 2).

Perfetti, C. A., Lindsey, R., & Garson, B. *Association and uncertainty: Norms of association to ambiguous words.* Pittsburgh: Learning Research and Development Center, University of Pittsburgh, 1971.

Random House dictionary of the English language. (College ed.) New York: Random House, 1968.

Schulman, A. I. Word length and rarity in recognition memory. *Psychonomic Science*, 1967, **9**, 211–212.

Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

Winograd, E., & Conn, C. P. Evidence from recognition memory for specific encoding of unmodified homographs. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 702–706.

(Received September 4, 1973)

Journal of ... Psychology
1974, Vo... 6, 1069–1075

EVI... OF A PRIMARY FRUSTRATION EFFECT FOLLOWING ...ALITY REDUCTION IN THE DOUBLE RUNWAY [1]

HENRY A. CROSS [2] AND WILLIAM N. BOYER

Colorado State University

...ts were trained with sucrose reward in both the midbox and the endbox of ...e double alley. Subsequently, *S*s were divided into groups which continued ... receive half of their trials under training conditions (nonfrustration trials) ...nd half ("frustration" trials) under either zero, alfalfa, regular, or sucrose ...ward. There was evidence of a primary frustration effect (FE) when reward ...uality was reduced in the midbox. Experiment II examined quality reduc...on in the midbox in a between-*S*s design and again there was an FE with ...ward quality reduction in the midbox. Results were interpreted as com...tible with the notion that an FE can be produced with incomplete reward ...duction in the midbox.

The a... ...rs (Boyer, Cross, & Anderson, in press) ...ve recently observed in both barpre... ...d free-choice preference situations ... ...s prefer, at least in limited eating sit...ons, Noyes 93% sucrose pellets to the re... ...ellet of the same manufacturer. ...s also prefer the regular pellet to the ...yes alfalfa-base pellet. The ordinal ...tionship of these reward preferences m... ...s possible a test for a primary frustrat... effect (FE) in a double-runway situation (A...sel, 1967) where midbox (Goal Box 1) reward can be "reduced," eve... though the weight or quantity of reward is actually held constant. According to Scull (1973), who has surveyed the frustration literature, there has never been an adequate demonstration that an FE can be produced through nonzero or incomplete reward reduction in the midbox of the traditional frustration apparatus. A number ... ...eople (e.g., Bower, 1962; Daly, 1968; McHose & Ludvigson, 1965) have tried to demonstrate the energizing effects of reward magnitude reduction in the midbox, but such attempts have either been que...tioned on methodological grounds or have raised some new problem. The present investigation is an attempt to circumvent some of the previous difficulties and to ascertain if incomplete reward reduction in Goal Box 1 leads to energized behavior in Alley 2 of the double runway.

[1] The authors wish to thank Barbara N. Hammack and Michael J. Sandifer who collected the data of Experiment I.

[2] Requests for reprints should be sent to Henry A. Cross, Department of Psychology, Colorado State University, Fort Collins, Colorado 80521.

Experiment I

Method

The *S*s consisted of 36 male albino rats from Carworth of New City, New York, which were 60–90 days old at the start of the experiment.

A traditional two-alley frustration apparatus was used throughout the experiment. All start and goal boxes were 30 cm. long, and the two alleys were 89 cm. in length. The apparatus was 15 cm. wide and 30 cm. in depth and was painted gray except for the black endbox (Goal Box 2). The floors of the start box, midbox (Goal Box 1), and endbox were spring loaded and permitted the recording of latency and running times in both segments. When *S* was placed in the initial start box, an electric circuit was opened which was later closed when *E* raised the start-box door. The first Standard electric clock, begun with the opening of the door, stopped when *S* left the start-box plate. Clock 2, which started with *S*'s departure from the start-box plate, stopped as *S* entered and depressed the midbox (Goal Box 1) plate. Clocks were read, reset, and readied for identical activation in the second alley. Times were recorded in hundredths of a second.

The *S*s were housed in pairs and allowed a 5-day habituation period with food and water ad lib. On Day 6, *S*s were placed on a 23-hr. deprivation schedule which continued throughout the experiment except for the time constraints encountered as the number of trials was increased during acquisition. On Days 8–10, *S*s were placed in pairs in the open apparatus and allowed 10 min. of exploratory activity. On Days 11–12, *S*s were allowed 10 min. of rewarded exploratory activity which consisted of free access to a cup of Noyes 45-mg. regular pellets in each goal box.

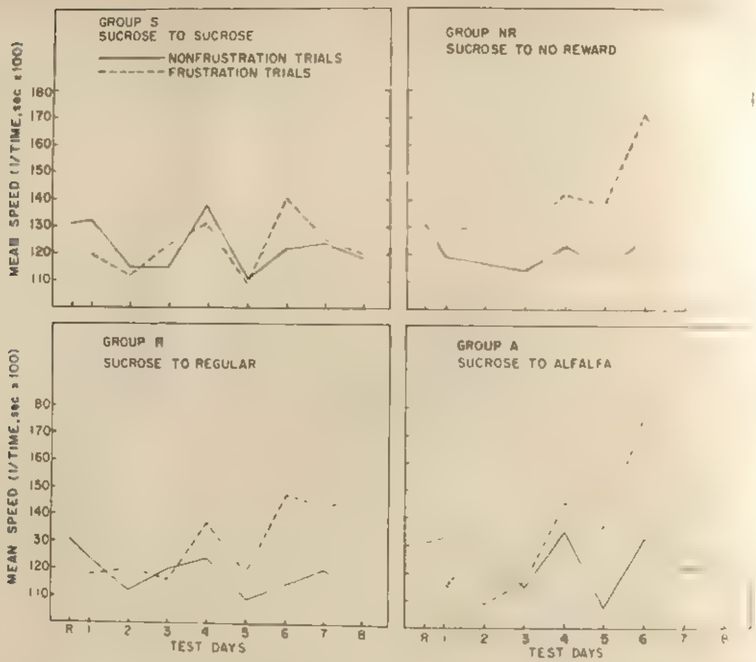


FIGURE 1. Alley 2 speed scores for frustration and nonfrustration test trials for the four basic groups over eight test days. (Abbreviation: R = mean score for the group at the end of acquisition training.)

Acquisition training was begun on Day 13 and continued through Day 35. All *Ss* were given a total of 66 trials which were rewarded in both the midbox and the endbox with four 45-mg. Noyes 93% sucrose pellets. The trials were initially given at the rate of 1 trial/day but were ultimately increased to 6 trials/day according to the following daily schedule: 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 6, 6, 6, 6, 6, 6, and 6. The *Ss* were given 3 sec. in the start box before the door was raised and were subsequently retained for 30 sec. in the midbox. The *Ss* were taken from the goal box after they finished eating their reward. All *Ss* were run in squads in a rotation which generated an intertrial interval of 10–15 min.

Following acquisition trials, the total running speed for Alley 2 was determined for the last six acquisition trials. All 36 *Ss* were rank ordered in terms of their terminal running speed and the four fastest *Ss* were then "blocked" and each member of the "block" was randomly assigned to a different treatment group. This procedure was continued until *Ss* had been distributed into nine such blocks. During the test trials, the four treatment groups each continued to receive two daily trials in which sucrose pellets, as in training, were available in both the midbox and the endbox. These daily test trials were ordered according to the following schedule of frustration (F) and nonfrustration (N) trials: FNFN, NFFN, FFNN, NFNF, FNNF, NNFF, FNFN, and NFFN. On the other two daily trials, the "frustration" trials, *Ss* received either sucrose reward, regular reward, alfalfa reward, or no reward in the midbox according to their respective group designation: S, R, A, or NR, respectively. When reward was given on designated frustration trials it always involved the same weight amount of food which had been employed during acquisition, i.e., four 45-mg. pellets. There was no change in endbox reward from that employed during acquisition. During this test phase, trials were run at the rate of 4/day and were continued for a total of eight days.

Results

The time scores (in seconds) for the second alley were reciprocally transformed (1/Time × 100) for each test trial. Subsequently, a mean speed score was computed for the frustration and nonfrustration conditions for each test day, and these mean speed scores served as the basic datum in a multifactored randomized block analysis of variance.

Figure 1 graphically depicts the experimental findings over the eight test days. It can be seen that in Group S there is no reliable difference between mean running speed on frustration and nonfrustration trials. This result was predicted since, in

this co[illegible] group, there is actually no differenti[illegible]atment for frustration and nonfrustra[illegible] trials. However, for the other groups [illegible]ere is a noticeable increase in runnin[illegible]peed as one moves from the nonfrustra[illegible]n to the frustration trials and this differe[illegible], i.e., frustration vs. nonfrustration, is [illegible]tatistically significant as indicated by b[illegible] a significant main effect for frustr[illegible]n condition, $F\ (1, 8) = 62.67$, $p <$ [illegible]05, and a significant Group × Frust[illegible]on Condition interaction, F (3, 324) [illegible]4, $p < .002$. There were two other [illegible]istically reliable differences: one differe[illegible]elated to an overall increased speed [illegible] test days, $F\ (7, 56) = 11.66$, $p <$ [illegible], and the other to a Days × Fru[illegible]ation Condition interaction, $F\ (7, 56) = 10.00$, $p < .000001$. These basic statis[illegible]al findin[illegible] revealed in Figure 1.

[illegible]omparisons with Duncan [illegible] indicated that only in [illegible] no reliable difference in m[illegible]peed for all shift trials between frustr[illegible]on ($\bar{X} = 122.89$) and nonfr[illegible] conditions ($\bar{X} = 121.97$). In the [illegible] groups the frustration tri[illegible]duced significantly faster [illegible] did the corresponding nonfrustration trials: Group R ($\bar{X} = 131.06$ vs. $\bar{X} = 117.00$), Group A ($\bar{X} = 138.39$ vs. $\bar{X} = 124.83$), and Group NR ($\bar{X} = 143.29$ vs. $\bar{X} = 117.49$).

[illegible]urther demonstration of the increased running speed produced by quality reduction in the midbox can be seen in the group comparisons on nonfrustration trials. The only significant comparisons were obtained between Group A and Group NR and between Group A and Group R. The slight elevation in running speed on nonfrustration trials in Group S animals was an anomalous result. In Groups S, R, and NR, the nonfrustration condition comparisons did not differ. A comparison of only the frustration condition across groups suggests something of the relative strength of a possible FE produced by quality reduction in the midbox and bears on the initial hypothesis that an FE would be inversely related to quality of reward obtaining on frustration trials. Groups R, A, and NR all differed significantly from the control situation, Group S, in frustration condition speeds. Beyond that, Group A did not differ from Group NR but did significantly exceed Group R and Group S in running speed. Group R *Ss* were also significantly faster than Group S animals on frustration trials.

Discussion

The results of this study suggest that it is possible to produce a primary FE by nonzero or incomplete reward reduction. In this particular case, the incomplete reward reduction was achieved, in Group A and Group R animals, while actually holding constant the weight of the food available on frustration and nonfrustration trials. The fact that *Ss* of Groups A and R had available on frustration trials a reward of equal weight and one that most probably contained more nutritional or satiety value than the nonfrustration sucrose reward would seem to militate against any attempt to invoke the original form of the "demotivation" or "response-depression" hypothesis (Seward, Pereboom, Butler, & Jones, 1957) as an explanation of the present results. On the other hand, it is true that the design of the present study does not effectively guard against the interpretation that the observed results are attributable to absolute effects stemming out of such things as subtle differences in eating times caused by different reward texture, etc. Another interpretation problem results from the fact that there was a Frustration Condition × Days Interaction. A perusal of Figure 1 shows that, with the exception of Group S, there is an increasing difference between frustration and nonfrustration trials over test days. It is possible that what is interpreted in the present study as an unlearned energizing effect of frustration is contaminated with an association component. For these reasons, it was decided that a second experiment should be attempted in which control conditions are established in a between-*Ss* design.

Experiment II

Method

The *Ss* were 44 male albino rats purchased from National Breeding Laboratories, Creve Couer, Missouri, which were 60–90 days old at the beginning of the experiment. The same two-alley frustration apparatus (Experiment I) was employed except that it was modified by painting the second alley black

and by placing ½-in. hardware cloth on the start box, first alley, and midbox. Both Alley 2 and the endbox (Goal Box 2) were black and were devoid of hardware cloth.

The Ss were housed in pairs and were allowed a 20-day habituation period in their cages with food and water ad lib. On Day 21, Ss were placed on a 23-hr. food deprivation schedule which continued throughout the experiment except for the time constraints encountered as the number of trials was increased during acquisition. On Days 22–26, Ss were placed, before their feeding period, in groups of six in the open apparatus devoid of all food cups and were allowed 5 min. of nonrewarded exploratory activity. On Days 27–29, the deprived Ss were placed in groups of two in an unpainted rectangular pine box (46 × 28 × 25 cm.) and were permitted to eat specific reward pellets from specific and differentiable food cups affixed on opposite ends of the box. The food cups were of different shape and the white one, which was later to occupy the midbox (Goal Box 1), was either empty or contained a large supply of 45-mg. pellets appropriate to the experimental group in which the two Ss had been randomly placed. The black cup, which was to occupy the endbox or Goal Box 2, always contained a large quantity of 45-mg. pellets appropriate to the experimental group in which the two Ss had been randomly placed. The black cup, which was to occupy the endbox or Goal Box 2, always contained a large quantity of 45-mg. 93% sucrose pellets since all Ss, regardless of their experimental group, were ultimately to receive such sucrose reward in the endbox. On Days 30–32, all 44 Ss were given, before their feeding period, a single goal placement in the black endbox of the two-alley frustration apparatus and were allowed to eat five 45-mg. Noyes pellets composed of 93% sucrose. The endbox door was down and the pellets were given in the black food cup appropriate to the endbox (Goal Box 2). On Day 33, all Ss received their first acquisition trial in the entire apparatus. The S was placed in the start box and the start-box latency and Alley 1 running times were recorded as in Experiment I. When S entered the midbox he received either no reward, 5 pellets of Noyes 45-mg. alfalfa pellets, or 93% sucrose. The difference in midbox reward constituted the only difference in the treatment of Ss during acquisition. After a 10-sec. delay in the midbox, E activated the midbox door and again latency and running times were recorded for the second portion of the apparatus. For 4 days, Days 33–36, each rat received only a single trial in the double-alley apparatus. This was followed by 6 days, Days 37–42, of 2 daily trials. On Day 43, the number of trials was increased to 3 and this schedule was continued through Day 63 when all Ss had received 79 trials in the double runway. On Day 64, the first frustration trial was given on the third daily trial, Trial 82, and these test days continued for a total of 12 days, Days 64–75.

At the start of the experiment all Ss were randomly assigned to one of four experimental groups ($n = 11$) which differed only in their midbox (Goal Box 1) reward experience. Group A-S Ss always received five 45-mg. alfalfa pellets in the midbox and five 45-mg. 93% sucrose pellets in the endbox. This reward experience was continuous over the entire experiment and did not change during the final 12-day test period. Group O-S Ss also received continuous training throughout. Group O-S animals differed from Group A-S Ss only in that on all trials they found no reward in the white food cup in the midbox. Group SS-A Ss all received 81 training trials in which both the midbox and the endbox always contained five 45-mg. 93% sucrose pellets. On Trial 82, the initial frustration trial, SS-A Ss received five 45-mg. alfalfa pellets instead of their customary 93% sucrose. Since the SS-A Ss had never before encountered alfalfa, and since pilot work indicated that the first encounter with alfalfa produces "shock" and long midbox latencies and Alley 2 running speeds, these Ss were given five alfalfa pellets to eat from a Gerber's baby jar lid in the neutral pine box approximately 1.5 hr. before their first test experience. This was done only once and was not repeated during the additional test days. The Ss of Group SS-O received five 45-mg. 93% sucrose pellets in both the midbox and endbox for the first 81 trials. Again, on Trial 82, these Ss received their initial frustration trial when they encountered an empty midbox. On the 12 test days, the initial daily trial for the two frustration groups, SS-A and SS-O, was always rewarded. The single daily frustration trial, when, occurred either on Trial 2 or Trial 3 and was randomly determined according to a schedule which had only the restraint that there be an equal number of frustration trials in each position (Trial 2 and Trial 3) over any 4-day test block. The actual schedule during test days was as follows: NNF, NFN, NFN, NNF, NFN, NFN, NNF, and NNF.

Results

To test that the four treatment groups were performing in a similar way at the end of acquisition, the Alley 2 time scores for the last six trials of acquisition (Trials 74–79) were reciprocally transformed and subjected to a split-plot analysis of variance. The group or treatment effect was not significant, $F\ (3, 40) = .31$, and the Group × Trials interaction was also insignificant, $F\ (15, 200) = .64$. Only the trials effect was significant, $F\ (5, 200) = 6.69$, and this was attributable to the fact that the initial daily trials (in this case Trial 74 and Trial 77) were depressed relative to the other trials within a given day.

Since all groups were essentially alike in Alley 2 speed at the end of acquisition, additional analyses centered around Alley 2 speed during the 12 test days in which

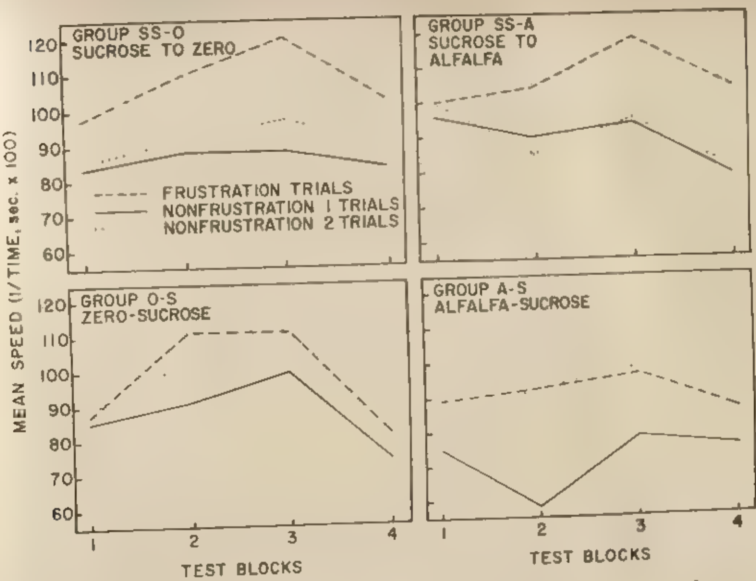


FIGURE 2. Alley 2 speed scores for Nonfrustration Trial 1, Nonfrustration Trial 2, and frustration trials for the four basic groups over four test blocks.

frustration trials were administered. Alley 2 runway times were reciprocally transformed and the transformed speed scores were collapsed into four blocks of three days each. The mean transformed value for a given block was then computed for the first nonfrustration trial, the second nonfrustration trial, and the frustration trial. These mean speed values were employed as the basic datum in a multifactored split-plot analysis of variance involving groups, frustration condition, and blocks. Figure 2 depicts the speed data.

The main effect for groups was not significant, F (3, 40) = .89, but there was a significant frustration condition effect, F (2, 80) = 28.42, $p < .000001$, and a significant first-order Groups $\times$ Frustration Condition interaction, F (6, 80) = 3.36, $p < .006$. In addition, the only other significant effect was that for blocks, F (3, 120) = 8.14, $p < .00006$.

Subsequent comparisons with the Duncan range test ($\alpha = .001$) revealed that all major comparisons were as predicted. The *Ss* of Group SS-A ran significantly faster on the frustration trials ($\bar{X} = 106.59$) than on either the first ($\bar{X} = 89.35$) or second ($\bar{X} = 90.05$) nonfrustration trial. In addition, these same *Ss* ran significantly faster on the frustration test trials ($\bar{X} = 106.59$) than the control *Ss*, Group A-S, ran on the same trials ($\bar{X} = 90.52$) but under conditions of continuous alfalfa reward in the midbox. The designation of a frustration trial for Group A-S animals was arbitrary and was dictated by the reward schedule of the parallel Group SS-A animals.

The other experimental group, Group SS-O, also showed the within-*Ss* FE in that mean frustration trials ($\bar{X} = 106.98$) significantly exceeded mean first ($\bar{X} = 85.50$) and second ($\bar{X} = 90.77$) nonfrustration trials. In addition, and more importantly, the frustration trial ($\bar{X} = 106.98$) of Group SS-O animals was significantly faster ($p < .025$) than the mean ($\bar{X} = 97.48$) of designated frustration trials in the corresponding control group, i.e., Group O-S.

As a test of the relative capacity of alfalfa test trials to energize behavior, i.e., produce the FE, it can be seen that the mean ($\bar{X} = 106.59$) of frustration trials of Group SS-A animals is essentially that of the mean ($\bar{X} = 106.98$) of the frustration trials of Group SS-O *Ss*. Neither experimental group differs significantly in their nonfrustration trials and the corre-

sponding frustration means in their respective control groups, Group A-S and Group O-S, are not reliably different.

The absence of any significant interaction involving blocks argues against the notion that the observed increase in speed on frustration trials is essentially acquired or is contaminated by an association component.

Finally, another observed result which was not anticipated was the greatly depressed initial daily trial, i.e., the first nonfrustration, of both of the control groups, Group A-S and Group O-S. In both groups, these means, 70.16 and 87.27, are significantly below their corresponding second nonfrustration trials, i.e., 90.57, 100.70. The initial nonfrustration trial for Group O-S animals does, however, appear to be essentially in line (compare with 89.35 and the 85.50 of the same trial for Group SS-A and Group SS-O). In Group O-S, then, it is the second nonfrustration trial which appears to be "out of line." At any rate, there is no ready explanation for this first-trial depression of Alley 2 speed in Group A-S and Group O-S *Ss*.

General Discussion

The reported experiments, particularly Experiment II, [illegible] that an FE is [illegible] expectancy for a higher, i.e., preferred, reward quality is not fulfilled. Such a notion is compatible with common [illegible] Tinklepaugh (1928) long ago demonstrated, at least at an anecdotal level, that a [illegible] with a normally [illegible] Bower (1962) sought to [illegible] nonreward occurs after reward [illegible] has been established [illegible] reduced to a lower [illegible] McHose and Ludvigson (1965) and [illegible] Peyser, and McHose (1965) cast doubt on Bower's interpretation by suggesting that his observed graded effects of nonzero reward reduction were co[...]ded by response inhibition stemming ou[t of] the rat's consumption of more or less foo[d in] the first goal box. The argument is that a [...] *Ss* design like Bower's does not prec[lude the] possibility that what appears to be a[n incre]ment in frustration drive with greate[r com]plete reward reduction is, in fact, si[mply a] function of absolute reward magnitude i[n Goal] Box 1. In short, the Bower results ar[e per]fectly compatible with the so-called "de[mo]tion hypothesis," which attributes diffe[rences] in Alley 2 running speed between frust[ration] and nonfrustration trials not to incr[eased] energy on frustration trials, but rather [to] depressed running speed following th[e re]warded nonfrustration trials.

Still later, Daly (1968) objected to cer[tain] inadequacies of the McHose and Ludvig[son] (1965) procedure. While these latter inve[sti]gators had, unlike Bower (1962), emplo[yed] control groups to test for a possible Alle[y] speed depression stemming out of a decre[ase] in hunger drive, they too encountered so[me] methodological difficulties. Even with th[e] controls, they did have their *Ss* experience more than one reward-reduction conditi[on], i.e., a partial within-*Ss* design. They also failed to control for the direction of shift in that only in one group, the group shifted to zero, was the shifted magnitude below the magnitude value employed in the second goal box. Daly's work pointed up further complexities in this area in that her attempt to deal with previous design limitations led in some groups, to differential acquisition performance in Alley 2 even though equal rewards were given in Goal Box 2.

The present approach sought to avoid some of these difficulties by employing a procedure wherein reward "reduction" in the midbox was restricted to a quality dimension. In Experiment II reward magnitude as such was always the same in both goal boxes, except for one group, Group SS-O. Although the previous design complexities in this area prompt caution, it is reasonable to conclude that the present results are compatible with the interpretation that an incomplete reward reduction in the midbox leads to a primary frustration effect as exhibited by faster running in Alley 2 on designated frustration trials. It could be maintained that the present findings result from quality change per se and cannot be unequivocally attributed to quality reduction or downward change. There are studies, for example, which show the invigorating effects of reward

chan[illegible]r se (Meyer & McHose, 1968; Stein[illegible] 1968), and such an interpretation cann[illegible] discounted. On the other hand, the [illegible]l changes in Experiment I which paral[illegible]e reward preferences argue against this i[illegible]rpretation. In addition, a recent replic[illegible]n by the authors[3] of the Crespi (1942) perf[illegible]ance shifts with reward quality change, in [illegible] of reward magnitude, offers further evid[illegible] against a simple change hypothesis.

[3] [illegible] Crespi, W. N. Boyer, and J. T. Melella. Un[illegible]ed study, 1973.

REFERENCES

Amsel, [illegible] Partial reinforcement effects on vigor [illegible] persistence. In J. T. Spence (Ed.), *The p[illegible]y of learning and motivation.* Vol. 1. N[illegible]k: Academic Press, 1967.

Bow[illegible] R. J., Peyser, C. S., & McHose, J. H. Effe[illegible]of complete and incomplete reward reduction o[illegible] a subsequent response. *Psychonomic Science,* 1965, 3, 277–278.

Bower, G. H. The influence of graded reductions in reward and prior frustrating events upon the magnitude of the frustration effect. *Journal of Comparative and Physiological Psychology,* 1962, 55, 582–587.

Boyer, W. N., Cross, H. A., & Anderson, C. Quality reward preference in the rat. *The Bulletin of the Psychonomic Society,* in press.

Crespi, L. P. Quantitative variation of incentive and performance of the white rat. *American Journal of Psychology,* 1942, 55, 467–517.

Daly, H. B. Excitatory and inhibitory effects of complete and incomplete reward reduction in the double runway. *Journal of Experimental Psychology,* 1968, 76, 430–438.

McHose, J. H., & Ludvigson, H. W. Role of reward magnitude and incomplete reduction of reward magnitude in the frustration effect. *Journal of Experimental Psychology,* 1965, 70, 490–495.

Meyer, P. A., & McHose, J. H. Facilitative effects of reward increase: An apparent "elation effect." *Psychonomic Science,* 1968, 13, 165–166.

Scull, J. W. The Amsel frustration effect: Interpretations and research. *Psychological Bulletin,* 1973, 79, 352–361.

Seward, J. P., Pereboom, A. C., Butler, B., & Jones, R. B. The role of prefeeding in an apparent frustration effect. *Journal of Experimental Psychology,* 1957, 54, 445–450.

[illegible], W. M. Response rate and varied reinforcement: Reinforcers of similar strength. *Psychonomic Science,* 1968, 10, 35–36.

Tinklepaugh, O. L. An experimental study of representative factors in monkeys. *Journal of Comparative Psychology,* 1928, 8, 197–236.

(Received October 12, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1076–1084

VARIABLES AFFECTING THE INTERMANUAL TRANSFER AND DECAY OF PRISM ADAPTATION[1]

CHONG SOOK CHOE[2] AND ROBERT B. WELCH

University of Kansas

The study examined the effects of exposure (continuous vs. terminal) and practice (massed vs. distributed) upon prism adaptation, its intermanual transfer, and spontaneous decay. Subjects were measured before and 0, 5, 10, and 15 min. after 30 trials of exposure to 11.3° lateral prism displacement on target-pointing accuracy (each limb separately), placing the arm straight ahead (each limb separately), and setting a pinpoint light source straight ahead. Adaptive pre-post shifts in these measures represented "negative aftereffect," "proprioceptive shift," and "visual shift," respectively. The major findings were that (*a*) negative aftereffect was greater with distributed than with massed practice, (*b*) all four groups experienced intermanual transfer of negative aftereffect, (*c*) proprioceptive shift transferred intermanually for the massed practice but not the distributed practice groups, (*d*) visual shift occurred only for the distributed practice–terminal exposure group, and (*e*) both ipsilateral and contralateral negative aftereffect and proprioceptive shift decayed significantly over the 15-min. postexposure period.

It is now well documented that human beings are capable of adapting to the effects of prismatically displaced vision (e.g., Hay & Pick, 1966; Held & Hein, 1958). Of recent interest is the issue of whether or not these effects transfer, in part or whole, from exposed to nonexposed limb. Some investigators (e.g., H. B. Cohen, 1963; Mikaelian, 1963) have reported no such transfer. However, Hamilton (1964) observed adaptive shifts for both exposed and nonexposed limbs when *S* was allowed free head movement during prism exposure; no transfer occurred when (as in previous studies) *S*'s head was immobilized. He suggested that the freedom to engage in head movements leads to a change in felt orientation of head relative to the remainder of the body, which in turn affects pointing localization for both hands.

The situation is complicated by the fact that a number of studies (e.g., Goldberg, Taub, & Berman, 1967; Kalil & Freedman, 1966) have demonstrated partial intermanual transfer of prism adaptation with *S*'s head in a fixed position during exposure. The crucial difference between experiments which have found intermanual transfer and those which have not may be in the nature of the exposure condition employed. Where intermanual transfer has not occurred *S* has been allowed a noninterrupted view of his limb throughout the prism-exposure period, while in experiments in which transfer was obtained, exposure to the limb occurred only at the termination of each visuomotor response. M. M. Cohen (1967) directly compared continuous and terminal prism exposure with respect to intermanual transfer of prism adaptation and his results lend support to this argument; only terminal exposure resulted in transfer. In an attempt to interpret this result, Taub and Goldberg (in press) have proposed that continuous exposure is actually a form of "massed" practice, whereas the terminal exposure condition represents "distributed" practice. They pointed out that Spatz (1966) had found intermanual transfer of a visuomotor skill (pursuit rotor) only when there was a rest interval between ipsilateral training and contralateral testing. Other investigators (e.g., Irion & Gustafson, 1952; Kimble, 1952) have reported similar results. Taub and

[1] The present investigation formed the basis of a master's thesis by Chong Sook Choe. The research was supported by Grant EY-00560-02 from the National Eye Institute of the National Institutes of Health. The authors wish to thank John J. Uhlarik and Robert I. Bermant for their comments and criticisms of this manuscript.

[2] Requests for reprints should be sent to Chong Sook Choe, Department of Psychology, University of Kansas, Lawrence, Kansas 66045.

Goldberg (in press) supported their hypothesis by finding intermanual transfer with a continuous exposure condition when it was interspersed with short rest intervals. Nevertheless, it remains tenable that both variables—type of exposure and type of practice—influence the occurrence and amount of intermanual transfer of prism adaptation. The primary goal of the present experiment was to investigate this possibility.

A second aim was to examine the nature of the adaptive shift underlying the intermanual transfer of adaptation. At least two end products of prism adaptation have been identified—a change in felt limb position (e.g., Harris, 1965) and a change in egocentric visual direction (e.g., Uhlarik & Canon, 1971). The latter investigators have hypothesized that the relative magnitudes of these two components of prism adaptation are determined by which informational source is attended to during prism exposure; it is primarily the nonattended modality that is modified. They argued that with terminal exposure *S* is forced to pay attention to the felt position of his limb throughout its movement, until finally seeing its displaced image. Thus, in this condition felt limb position should remain relatively unmodified, while vision is changed. Conversely, vision "dominates" in the continuous exposure condition and therefore proprioception should be altered. Their results confirmed these predictions. Therefore, it is conceivable that the presence of intermanual transfer of prism adaptation only in experiments using terminal exposure results from the fact that this condition elicits a shift in vision, an effect which necessarily determines pointing accuracy for both limbs. Clearly, this interpretation differs from that suggested by Taub and Goldberg (in press). Finally, there remains the possibility that intermanual transfer of a shift in felt limb position also occurs and represents at least a partial basis for intermanual transfer of the prism-corrective changes in eye–hand coordination.

A third goal of the present investigation was to examine the spontaneous decay of prism adaptation, in both exposed and nonexposed limbs, as a function of exposure condition and type of practice. Some investigators (Hamilton, 1964; Hamilton & Bossom, 1964; Taub & Goldberg, in press) have reported rapid decay of negative aftereffect during postexposure periods of 15 min. in duration. Others (e.g., Dewar, 1970; Goldberg, Taub, & Berman, 1967; Welch, 1971) failed to observe decay during this same postexposure interval. Again, it appears that exposure condition is an important variable. All the studies employing continuous exposure have produced rapid decay; those involving terminal exposure have led to little or no decay. It was possible to directly compare these two conditions in the present experiment. Only two reported studies have examined the effect of type of practice upon decay of adaptation, and their findings appear contradictory. Dewar (1970), using terminal exposure, found that adaptation resulting from massed practice decayed faster than adaptation generated in a distributed practice condition. On the other hand, Taub & Goldberg (in press), who used continuous exposure, appear to have found just the opposite result. The implication that type of exposure and type of practice interact with respect to rate of prism adaptation decay was examined in the present study.

METHOD

Design

All *S*s were measured before and after prism exposure on accuracy of target pointing (visuomotor coordination) and placing the arm straight ahead of the nose (felt limb position). For each task both exposed and nonexposed hands were tested. A third measure was *S*'s accuracy in setting a pinpoint light source straight ahead of his nose (visual straight ahead). The postexposure measures of each of these three tasks were taken at 0, 5, 10, and 15 min. after the termination of the prism-exposure period. For none of the pre- or postexposure measures was *S* allowed visual feedback regarding the accuracy of his performance. Type of exposure—continuous and terminal—was combined with type of practice—massed and distributed—to produce four experimental groups.

Subjects

Sixty-four students from introductory psychology classes at the University of Kansas served as *S*s

in the experiment in order to fulfill a course requirement, All had normal vision without correction and were naive as to the nature of the experiment. They were randomly assigned to four groups, each containing 16 *S*s.

Apparatus and Procedure

Displacement of the visual field was produced by a Risley rotating prism (Bausch and Lomb, No. 71-48-59), mounted in the right eyepiece of a pair of welder's safety goggles. The left eyepiece of the goggles was occluded.

The main testing apparatus was patterned directly after one designed by Uhlarik (1972). The *S* sat in front of a table upon which was placed a horizontal board mounted 43 cm. above the surface of the table. A quarter circle was cut from the board and fitted with transparent Plexiglas. This "exposure slot" could be covered with a section of black posterboard during certain parts of the experiment, to keep *S* from viewing his limb. At the far edge of the exposure slot was a vertical, black pegboard panel, circumscribing an arc whose radius was 57 cm. (centered at *S*'s right eye). The panel extended approximately 11 cm. above and below the horizontal surface. A position transducer (resistance coil), 6 mm. in diameter and 90 cm. long, extended horizontally along the inner surface of the pegboard panel, 6 cm. below the exposure slot. On each of his index fingers *S* wore a rubber finger with a metal probe attached to it. The measure of target-pointing accuracy was taken by requesting *S* to reach forward along the underside of the covered exposure slot and touch the position transducer with the tip of the probe at a point directly beneath the apparent position of the target. The target was a pinpoint of light positioned 2.5 cm. above the surface of the slot. A digital voltmeter, which was located out of *S*'s view, registered the voltage at the point at which the probe made contact with the transducer, thus providing *E* with a measure of *S*'s target-pointing accuracy. The experimental session was divided into three phases: Preexposure, Exposure and Postexposure.

Preexposure Phase. At the outset of the Preexposure Phase an impression of *S*s' teeth was taken to provide the basis for a biteboard used during certain parts of the experiment to stabilize the position of his head. After instructions and several practice target-pointing trials (with visual feedback) the goggles were placed over *S*'s eyes and the variable prism set at 0 diopters (no displacement). Then *S* was required to point directly beneath the light, whose position varied nonsystematically among three different locations—straight ahead, 10 cm. to the left, and 10 cm. to the right. The exposure slot was covered, to preclude visual feedback to *S*. Three responses were made at each of the three target positions with one hand and then this was repeated with the other hand. The order of the target positions was the same for both limbs and for all *S*s. For one half of the *S*s the right hand was tested first and the left hand second; for the remainder the order was reversed. For one half of the *S*s the hand tested first was to be the prism-exposed one; the remainder were tested first with the nonexposed hand. For the measure of felt limb position *S* placed his arm so that it was felt to be located directly in front of his nose. The arm movement was similar to that used for the target-pointing response. Vision was occluded and accuracy was measured four times with each limb, the order of hand tested being the same as for the target-pointing response. The third type of preexposure measure was *S*'s accuracy in setting a pinpoint light source to the straight-ahead position in the otherwise dark room. The *S* pushed a toggle switch (using the hand that would not be exposed to the prism) which caused a dot of light to move laterally at a slow, constant speed (3.57 [illegible]/sec) just above the top of the pegboard panel, at approximately eye level. This light (T1-5 v., run at 1.5 v.) was attached to the end of a rod which projected vertically from a toy car located on an arced track behind the vertical pegboard panel. Four measures were taken, the starting position varying from far right to far left (in a "RLLR" order). The *S* was allowed to reverse the direction of the light if it had gone past apparent straight ahead. A tape recording of white noise was played during this measure in order to mask potential auditory location cues resulting from the sound of the toy car's electric motor. Preexposure target-pointing accuracy was always recorded first. However, measures of felt limb position and visual straight ahead were taken second and third, respectively, for half of the *S*s, and in the reverse order for the other half.

Exposure Phase. During the Exposure [illegible] of the experiment *S* was allowed to see his li[illegible] the exposure slot, which was either co[illegible] partially uncovered, depending upon the experimental condition. A luminous rubber finger was placed over the index finger of the hand to be exposed. The task during this part of the experiment was to reach forward and point to one of three targets—the numbers "1," "2," and "3" constructed of luminous tape and placed at the far edge of the board at points corresponding to the target positions used during the Preexposure Phase. With *S*'s eyes closed, the variable prism was adjusted to 20 diopters, with the base oriented either right or left, producing a lateral visual displacement of approximately 11.3°. The *S* was informed by *E* of the presence and nature of the visual rearrangement. After the room lights had been extinguished, *S* opened his eyes and repeatedly pointed at each of the three luminous numbers as they were called out in a predetermined block-randomized number sequence. The sequence was the same for all *S*s. The luminous rubber finger which *S* wore allowed him to see his target-pointing accuracy. A constant rate of responding was maintained by means of a tape-recorded 2-sec. metronome beat. The *S*'s hand began from a starting position at the near edge of the table and traveled approximately 60 cm. to its destination. Thirty unrecorded prism-exposure responses were made, 10 for each of the three targets.

Th[illegible] instructed to point as accurately as he co[illegible] ot to correct for an error while the po[illegible] ger was in view. It was during the E[illegible] hase that the experimental conditions w[illegible] mented.

[illegible] the Ss were provided with continuous ex- p[illegible] to the limb by completely uncovering the ex[illegible] re slot. The remainder experienced terminal e[illegible], for these Ss the slot was covered except fo[illegible] 3-cm. gap at its perimeter. These two groups w[illegible] further subdivided into massed and distributed [illegible] tice conditions. Massed practice was achieved [illegible] aving S point at a target on every 2-sec. beat [illegible] he tape-recorded metronome for the 30 exposure [illegible] ds, thus providing 1 min. of uninterrupted [illegible] ctice. In the distributed practice condition S [illegible] ponded every 2 sec., and then "rested" for 2 sec., [illegible] a total exposure period of 2 min. (i.e., 1 min. of [illegible] ctice). Thus, as defined by the nature of their [illegible] tivity during the Exposure Phase, there were four [illegible] roups: Continuous-Massed, Continuous-Distrib- [illegible] ted, Terminal-Massed, and Terminal-Distributed.

Postexposure Phase. Immediately after the final prism-exposure trial, the Postexposure Phase began. The S closed his eyes, the exposure slot was covered, the number targets removed, the variable prism reset to 0 diopters, and the luminous finger removed. The S was told that his vision was once again normal.[3] Then no-feedback measures were taken of target-pointing accuracy (three trials for each limb), felt limb position (two trials for each limb), and visual straight ahead (two trials). These three measures constituted the 0-min. Postexposure Test. The order of hand tested was the same as in the Preexposure Phase for a given S, as was the order in which the measures of visual straight ahead and felt limb position occurred. After these initial postexposure measurements had been obtained S was requested to rest quietly in his chair with eyes closed and arms placed at his sides. After 5 min., another set of measures was taken, (the 5-min. Postexposure Test). This procedure was repeated at 10 and 15 postexposure min. Including the time spent in taking the measures, the total length of the Postexposure Phase was approximately 25 min.

[3] The rationale for informing S that his vision was or was not displaced is that in prism exposure involving target pointing, and hence error-corrective feedback, the existence and nature of the visual rearrangement quickly becomes apparent to S. Consequently, it is crucial that he be informed prior to the Postexposure Phase that his vision will once again be normal. Otherwise, his negative aftereffect is likely to be contaminated by the conscious and deliberate correction for errors that he is likely to have used during the Exposure Phase and which he would continue to use if he thought that his vision remained displaced. It is assumed that the aftereffect of prism exposure is, at least with respect to target-pointing, a more valid measure of adaptation in the present experimental situation than is any change in behavior observed to occur during the Exposure Phase.

Results

Three operationally different measures of adaptation were obtained, each defined as the difference between pre- and postexposure performance. A prism-compensatory shift in target-pointing accuracy is commonly referred to as "negative aftereffect" (NA); a shift in felt limb position may be termed "proprioceptive shift" (PS) and a change in visual straight ahead, "visual shift" (VS). For the purposes of the statistical analyses, a pre-post difference in the adaptive direction was given a positive sign while an "antiadaptive" shift was negatively signed. Because magnitude of NA did not differ between the prism base-right and prism base-left conditions, $t\ (62) = .99$, $p > .05$, the data for the two prism orientations were combined for the subsequent analyses. Also, the data were collapsed over hand exposed, order of hand tested, and order of visual and proprioceptive measures. Figure 1 depicts the levels of the three measures of prism adaptation and their decay rates over 15 min. for each of the four experimental groups.

Negative Aftereffect. Clearly, all groups manifested with the exposed limb a mean NA (see Figure 1A), which was significantly greater than zero at the 0-min. Postexposure Test. With respect to the unexposed limb a t test revealed that the adaptive shift at the 0-min. test for all conditions combined was statistically significant, $t\ (63) = 6.19$, $p < .001$, indicating that partial intermanual transfer of adaptation had occurred. A three-factor mixed analysis of variance (Exposure × Practice × Limb) was carried out upon the NA scores at the 0-min. Postexposure Test. The results indicated significant effects for Practice, $F\ (1, 60) = 4.98$, $p < .05$; Limb, $F\ (1, 60) = 100.20$, $p < .001$; and their interaction, $F\ (1, 60) = 5.36$, $p < .025$. It is clear from an examination of Figure 1A that the interaction represented the fact that the difference in NA between massed and distributed practice (in favor of the latter) existed only for the exposed limb.

As a means of testing for the presence of decay of NA in the exposed limb across the four Postexposure Tests, a three-factor

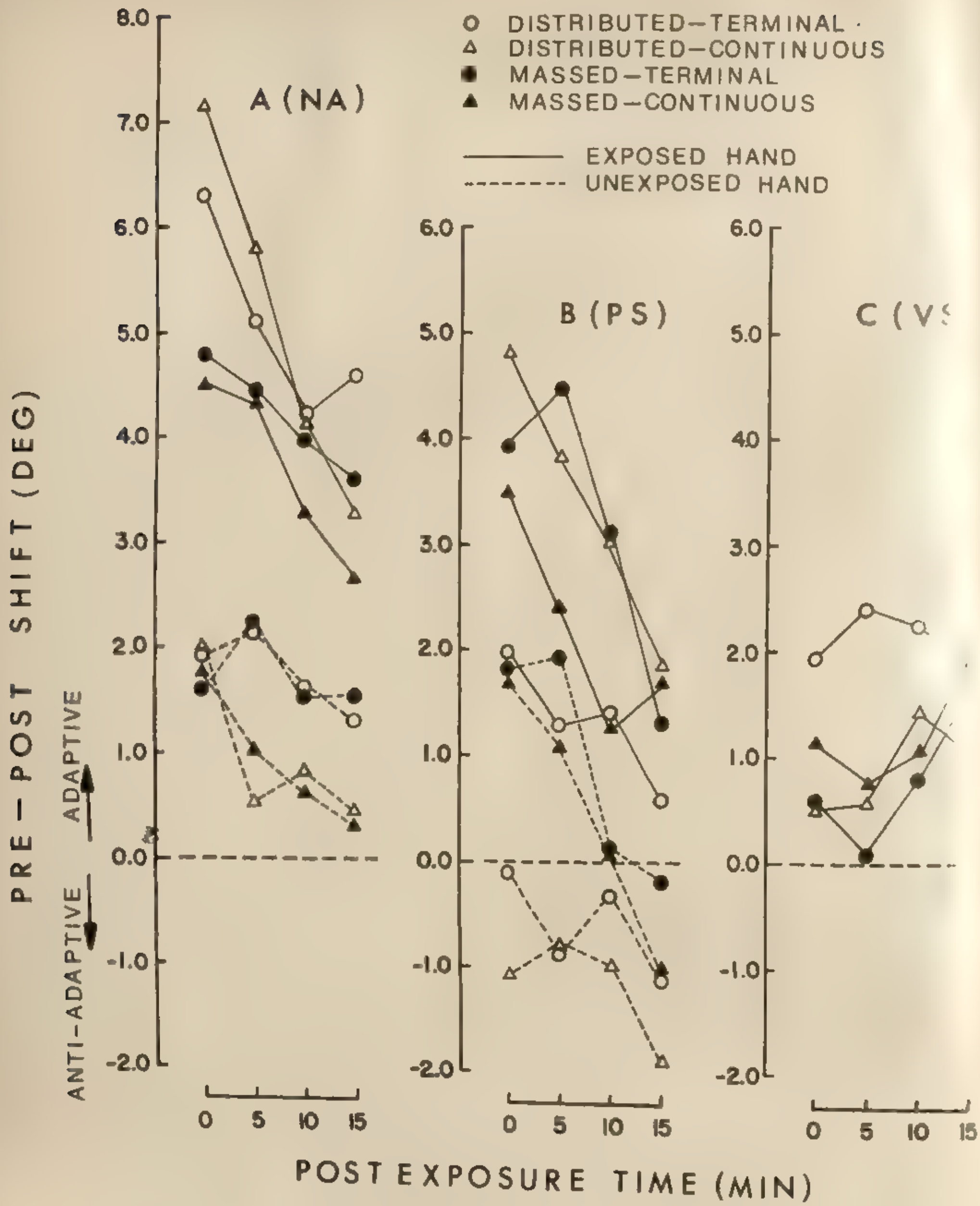


FIGURE 1. Mean Pre-Post Shifts (deg.) at each Postexposure Test. (Abbreviations: NA = Negative Aftereffect, PS = Proprioceptive Shift, VS = Visual Shift.)

mixed analysis of variance (Practice × Exposure × Postexposure Test) was performed on the data. The Postexposure Test effect was clearly significant, F (3, 180) = 20.81, $p < .001$, indicating that a certain amount of decay occurred. The interactions between Practice and Postexposure Test, F (3, 180) = 2.56, $p < .055$, and between Exposure and Postexposure Test, F (3, 180) = 2.59, $p < .055$, were marginally significant. The same analysis carried out on the results for the unexposed limb also revealed a significant F for Postexposure Test, F (3, 180) = 3.58, $p < .025$.

However, none of the remaining factors or interactions was statistically significant.

Proprioceptive Shift. Although it is apparent from inspection of Figure 1B that PS occurred for the exposed limb at the 0-min. Postexposure Test, it is not clear that this was true for the nonexposed limb. In fact, a t test on the PS scores for the nonexposed limb for all groups combined failed to find the mean to be significantly greater than zero, $t\ (63) = 1.15$. However, when each group was tested individually for contralateral PS the means for the terminal-massed and continuous-massed groups each proved to be greater than zero, $t\ (15) = 1.86$, $p < .05$; $t\ (15) = 1.93$, $p < .05$, indicating the presence of intermanual transfer of PS for these groups. A mixed analysis of variance (Exposure × Practice × Limb) on the PS measure for both limbs at the 0-min. Postexposure Test found only the difference between limbs to be statistically significant, $F\ (1,\ 60) = 27.00$, $p < .001$. A Practice × Exposure analysis of variance was carried out upon the PS measures for the nonexposed limb alone and revealed a significant effect for Practice, $F\ (1,\ 60) = 5.71$, $p < .025$ supporting the conclusion that only the massed practice groups experienced intermanual transfer of PS.

An Exposure × Practice × Postexposure Test analysis of variance on PS for the exposed limb revealed that the effect of Postexposure Test was statistically significant, $F\ (3,\ 180) = 8.43$, $p < .001$. This indicates that a decline in ipsilateral PS occurred during the postexposure period. None of the other factors or interactions was significant. The same analysis applied to the results for the nonexposed limb produced similar results—Postexposure Test was the only significant factor, $F\ (3,\ 180) = 3.49$, $p < .025$.

Visual Shift. A t test revealed that mean VS at the 0-min. Postexposure Test for all groups combined was significantly greater than zero, $t\ (63) = 2.60$, $p < .01$. However, individual t tests revealed that only the mean for the terminal-distributed group was statistically significant, $t\ (15) = 2.06$, $p < .05$. A two-factor analysis of variance (Exposure × Practice) on the data for the 0-min. Postexposure Test revealed no significant main effect nor an interaction. When Postexposure Test was added to these factors, in order to examine potential changes of VS over the Postexposure Phase, none of the factors or interactions was significant.

Comparison of Negative Aftereffect, Proprioceptive Shift, and Visual Shift. To examine the effects of practice and exposure on intermanual transfer of NA and PS a four-way mixed analysis of variance (Exposure × Practice × Type of Adaptation × Limb) was carried out on the 0-min. Postexposure Test measures. A significant F was obtained for Type of Adaptation, $F\ (1,\ 60) = 10.44$, $p < .005$. Comparison of Figure 1, A and B, reveals that NA was greater than PS. Limb was another significant factor, $F\ (1,\ 60) = 86.49$, $p < .001$, confirming the results of the analyses reported previously. Also statistically significant were the Practice × Type of Adaptation interaction, $F\ (1,\ 60) = 6.15$, $p < .025$, as well as the Practice × Limb interaction, $F\ (1,\ 60) = 7.02$, $p < .025$. Inspection of Figure 1 indicates that these interactions represented the fact that practice had a differential effect on NA (in favor of distributed practice) for the exposed limb only and on PS (in favor of massed practice) for the nonexposed limb only.

As a means of investigating the possibility that type of practice and type of exposure have differential effects on adaptation as a function of which measure is used, an Exposure × Practice × Type of Adaptation analysis was performed upon VS and ipsilateral PS taken at the 0-min. Postexposure Test. Only Type of Adaptation proved to be statistically significant, $F\ (1,\ 60) = 8.73$, $p < .005$. Examination of Figure 1, B and C, indicates that ipsilateral PS was greater than VS.

In order to examine the relative rates of decay of adaptation for each of the three measures, an analysis (Exposure × Practice × Type of Adaptation × Postexposure Test) was done on VS and the NA and PS measures taken with the exposed limb only. Type of Adaptation, $F\ (2,\ 120) =$

9.66, $p < .001$, Postexposure Test, F (3, 180) = 14.08, $p < .001$, and their interaction, F (6, 360) = 6.28, $p < .001$, all were significant. The interaction appears to be due to the fact that only VS failed to decay over the Postexposure Tests.

Discussion

Contrary to expectation, all four experimental groups demonstrated intermanual transfer of NA. It is possible that the relatively short prism-exposure period (30 trials of target pointing, spread over 1–2 min.) may have been an optimal situation for generating transfer, irrespective of experimental condition. Freedman (1968) reported large intermanual transfer when training was terminated at a preset criterion, while no transfer was observed when *S*s were overtrained. In the present experiment, intermanual transfer of NA was not influenced by type of exposure. This result is contrary to the differential exposure effect on transfer reported by M. M. Cohen (1967). It is not clear whether massed vs. distributed practice affected transfer. Initially, the absence of a main effect for Practice might be viewed as an indication that the type of practice did not affect the intermanual transfer of NA. However, intermanual transfer is traditionally evaluated in terms of the proportion of adaptation in the unexposed limb to that in the exposed limb. Defining transfer in this manner leads to an alternative interpretation; that is, different types of practice do indeed lead to different amounts of transfer. The significant Practice × Limb interaction of NA measured at the 0-min. Postexposure Test indicates that massed practice produced more transfer than distributed practice, and thus supports the second conclusion. At present, it is difficult to determine which of these interpretations is the most tenable. To clear up this problem an additional study is needed which would control the level of adaptation in the exposed limb. Thus, for instance, if a similar Practice × Limb interaction emerged under these conditions, it would support the second conclusion.

Intermanual transfer of PS occurred only for the two massed practice groups. A transfer of modified proprioception from exposed to nonexposed limb has never previously been reported and it is not at all apparent how this might occur. Another finding difficult to interpret is that in the two distributed groups the PS measure for the nonexposed hand showed the change in the antiadapti[illegible] direction. Perhaps the task of pointing [illegible]aight ahead of the nose in the dark is gove[illegible]d not only by felt limb position but also by a [illegible]ponse tendency to point at an imagined [illegible]ight-ahead position. It could be argued [illegible]ther that both proprioception and this [illegible]nse tendency are altered by prism expos[illegible] but that only the latter component (or a [illegible]tion of it) transfers to the contralater[illegible]mb. Aside from the validity of this explan[illegible] it remains of interest that only the [illegible]sed practice groups demonstrated transfe[illegible] PS. Furthermore, regardless of the basis [illegible] this form of transfer, its existence sugge[illegible]hat the intermanual transfer of NA m[illegible] not always be the result of a change in v[illegible]n, as some investigators (e.g., M. M. Cohen, [illegible]967; Kalil & Freedman, 1966; Wilkinso[illegible]71) have apparently assumed.

Both NA and PS revealed a signifi[illegible] decline across the four Postexposure T[illegible] for both limbs. A review of previous lite[illegible]re led to the prediction that NA (for ex[illegible]ed limb) resulting from continuous [illegible]re would decay more rapidly than that [illegible]d by terminal exposure. Such an Exp[illegible] Postexposure Test interaction was fou[illegible]n the present study, although the signifi[illegible]nce level was marginal ($p < .055$). One ex[illegible]nation for this tentative finding is that term[illegible]al prism exposure leads to a relativ[illegible]ge change in vision (cf. Uhlarik & Cano[illegible]) and that a prism-adaptive change i[illegible] visi[illegible] decays less rapidly than does propri[illegible] adaptation. Partial support for this [illegible]pretation is seen in the finding that VS fa[illegible]d to decay during the Postexposure Phase. In fact, there was a nonsignificant *increase* in VS for three of the four groups. The fact t[illegible]t VS appears to be more resistant to decay than PS may be related to the results of a study by Hay and Pick (1966) who found that ove[illegible] 6-day prism exposure period visual adaptation ultimately replaced the relatively transient initial change in felt limb position.

The results of the present investigation help to define the optimal conditions for inducing the various end products of prism adaptation. As already indicated, distributed prism exposure (whether terminal or continuous) led to significantly greater NA than did massed practice, a finding previously reported by Taub and Goldberg (in press) with continuous exposure, but not found by Dewar (1970), who used terminal exposure. It is possible that Dewar's use of the "prism shaping"

techn... of Howard (1968) led to such a lar... ount of adaptation for both groups tha... ffects of the practice variable were "ov... elmed." Examination of Figure 1B rev... an unclear picture regarding the op... al condition for eliciting PS. None of the ...actors or interactions for PS in the ex-p... d limb at the 0-min. Postexposure Test ...ed statistically significant. With respect ...S the most effective condition at the 0-min. ...texposure Test was distributed practice-...minal exposure. The finding by Uhlarik ...l Canon (1971) that terminal exposure ...duces primarily VS and continuous exposure ...marily PS was not supported by the present ...ata when the practice variable was collapsed. ...n the other hand, Figure 1, B and C, suggests ...he existence of this predicted Exposure × ...ype of Adaptation interaction, provided that only the distributed practice groups are examined. It should also be noted that in the study by Uhlarik and Canon (1971) the obtained interaction held only for the "direct" measures (i.e., those taken with the prism displacement present); no interaction was found for the aftereffect measures which, of course, were the only kind taken in the present experiment.

With the present data it is possible to examine the "additive model" of prism adaptation presented by several investigators (e.g., Wilkinson, 1971), which holds that NA = VS + PS. When adaptation as manifested at the 0-min. Postexposure Test was examined for all groups combined, PS + VS = 4.57, while NA = 5.64. This difference of 1.07 proved to be significantly greater than zero, t (63) = 1.74, $p < .05$. While it is possible that the finding of NA > PS + VS in this experiment was due to the fact that NA was measured before PS and VS and, therefore, was perhaps less affected by spontaneous decay, this seems an unlikely event, given that only three measures of NA for each hand (a total period of approximately 45 sec.) were taken prior to the measures of PS or VS. Furthermore, an interesting observation is made when this additive model is tested for each group separately. For the massed practice groups the model held quite well—the difference between NA (4.60) and PS + VS (4.55) was not significantly greater than zero, t (31) = .10, $p > .05$. On the other hand, the model failed to predict the results for the distributed practice groups, in that the difference between NA (6.68) and PS + VS (4.58) was different from zero, t (31) = 2.76, $p < .003$. The tentative conclusion is that distributed prism exposure leads to a third component of prism adaptation, which adds to PS and VS in determining the magnitude of NA. The precise nature of this postulated component remains to be clarified.

REFERENCES

Cohen, H. B. Transfer and dissipation of aftereffects due to displacement of visual field. *American Psychologist*, 1963, **18**, 411. (Abstract)

Cohen, M. M. Continuous versus terminal visual feedback in prism aftereffects. *Perceptual and Motor Skills*, 1967, **24**, 1295–1302.

Dewar, R. Adaptation to displaced vision—the influence of distribution of practice on retention. *Perception & Psychophysics*, 1970, **8**, 33–34.

Freedman, S. J. Perceptual compensation and learning. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior*. Homewood, Illinois: Dorsey Press, 1968.

Goldberg, I. A., Taub, E., & Berman, A. J. Decay of prism aftereffect and interlimb transfer of adaptation. Paper presented at the meeting of the Eastern Psychological Association, Boston, April 1967.

Hamilton, C. R. Intermanual transfer of adaptation to prisms. *American Journal of Psychology*, 1964, **77**, 457–462.

Hamilton, C. R., & Bossom, J. Decay of prism aftereffects. *Journal of Experimental Psychology*, 1964, **67**, 148–150.

Harris, C. S. Perceptual adaptation to inverted, reversed, and displaced vision. *Psychological Review*, 1965, **72**, 419–444.

Hay, J. C., & Pick, H. L., Jr. Visual and proprioceptive adaptation to optical displacement of the visual stimulus. *Journal of Experimental Psychology*, 1966, **71**, 150–158.

Held, R., & Hein, A. Adaptation to disarranged hand-eye coordination contingent upon re-afferent stimulation. *Perceptual and Motor Skills*, 1958, **8**, 87–90.

Howard, I. P. Displacing the optical array. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior*. Homewood, Illinois: Dorsey Press, 1968.

Irion, A. L., & Gustafson, L. M. "Reminiscence" in bilateral transfer. *Journal of Experimental Psychology*, 1952, **43**, 321–323.

Kalil, R. E., & Freedman, S. J. Intermanual transfer of compensation for displaced vision. *Perceptual and Motor Skills*, 1966, **22**, 123–126.

Kimble, G. A. Transfer of work inhibition in motor learning. *Journal of Experimental Psychology*, 1952, **43**, 391–392.

Mikaelian, H. H. Failure of bilateral transfer in modified eye-hand coordination. Paper presented at the meeting of the Eastern Psychological Association, New York, April 1963.

SPATZ, K. C. An analysis of bilateral and unilateral reminiscence. Unpublished doctoral dissertation, Tulane University, 1966.

TAUB, E., & GOLDBERG, I. A. Prism adaptation: Control of intermanual transfer by distribution of practice. *Science*, in press.

UHLARIK, J. J. A device for presenting targets and recording positioning responses in one dimension. *Behavior Research Methods and Instrumentation*, 1972, **4**, 15–16.

UHLARIK, J. J., & CANON, L. K. Influence of concurrent and terminal exposure conditions on the nature of perceptual adaptation. *Journal of Experimental Psychology*, 1971, **91**, 233–2[illegible].

WELCH, R. B. Prism adaptation: The "target-pointing effect" as a function of exposure trials. *Perception & Psychophysics*, 1971, **9**, 102–[illegible].

WILKINSON, D. A. Visual-motor control loop: A linear system? *Journal of Experimental Psychology*, 1971, **89**, 250–257.

(Received October 19, 1973)

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Memory for Modality: Within-Modality Discrimination is Not Automatic. Dale E. Berger (Faculty in Psychology, Claremont Graduate School, Benezet Psychology Building, Claremont, California 91711) and Leah L. Light.

Role of Olfactory Cues in the Acquisition and Extinction of Avoidance. J. K. Dua (Department of Psychology, University of New England, Armidale, New South Wales, Australia) and M. J. Dobson.

Retroactive Inhibition in a Bilingual A B, A′ B′ Paradigm. Robert K. Young (Department of Psychology, University of Texas at Austin, Mezes Hall 330, Austin, Texas 78712), Mike Lopez, and Robert E. [illegible]

Partial Repetition of Digit Strings with Increased Degree of Learning. James D. Fritzen (Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061).

Adaptation to Displaced Hearing: A Nonproprioceptive Change. H. H. Mikaelian (Department of Psychology, Connecticut College, New London, Connecticut 06320).

Paired-Associate Acquisition as a Function of Number of Initial Nontest Trials. James F. Voss (Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260) and Ronald [illegible] Porte[illegible].

Precategorical Selective Attention and Tonal Specificity in Auditory Recognition. Harold L. [illegible] (Department of Psychology, University of Oregon, Eugene, Oregon 97403), Gerald B. Thom[illegible] C. Presson, Andrew Cozic, and David Brookmire.

Frequency of Feedback and Learned Heart-Rate Control. Robert J. Gatchel (Department of Psychology, University of Texas at Arlington, Room 309 Life Science Building, Arlington, Texas 76010).

Free Recall as a Function of Type of Evoking Stimulus. Wilma A. Winnick (Department of Psychology, Queens College, City University of New York, Flushing, New York 11367), Fae Kooper, and Joyce Sprafkin.

Temporal Differentiation and Recognition Memory for Visual Stimuli in Rhesus Monkeys. Martha Wilson (Department of Psychology, University of Connecticut, Storrs, Connecticut 06268) and Mildred Mason.

Organization in Normal and Retarded Children: Temporal Aspects of Storage and Retrieval. Mark H. Ashcraft (Department of Psychology, University of Kansas, Lawrence, Kansas 66044) and George Kellas.

Positive Recency in Final Free Recall. Gregory F. Mazuryk (Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1M1).

The Prediction Outcome of a Response-Irrelevant Stimulus as a Determinant of Choice Reaction Time. E. Scott Geller (Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061).

Time Release from Proactive Inhibition as a Function of the Amount of Proactive Inhibition Present. David S. Gorfein (Department of Psychology, Division of Natural Sciences, New College, Sarasota, Florida 33578).

Response Rate and Development of UCS-Attenuating Response Topography in Eyelid Conditioning under Different Conditions of Reinforcement. David A. Grant (Department of Psychology, Psychology Building, University of Wisconsin, Charter at Johnson, Madison, Wisconsin 53706) and Joseph B. Hellige.

Detectability of Relative Motion as a Function of Exposure Duration, Angular Separation, and Background. Lewis O. Harvey, Jr. (Massachusetts College of Optometry, 424 Beacon Street, Boston, Massachusetts 02115) and John A. Michon.

Effects of Priority Instructions on the Processing of Hypercapacity Sequential Inputs of Pictures and Words. Matt Erdelyi (Department of Psychology, Brooklyn College, City University of New York, Brooklyn, New York 11210) and Lorraine F. Kubicek.

(*Continued on page 1091*)

Journa[illegible] [illegible]mental Psychology
1974, V[illegible] [illegible], No. 6, 1085 1091

[illegible]TRIEVAL OF WORDS FROM WELL-LEARNED SETS: THE EFFECT OF CATEGORY SIZE [1]

[illegible]ONATHAN L. FREEDMAN [2]
Columbia University

ELIZABETH F. LOFTUS
University of Washington

In four experiments, *Ss* learned categories of varying size and were then required to produce a member of one of the categories that began with a particular letter. There was only one word in the category that began with the letter. Reaction times increased as a function of size with small categories (less than 7 words) but at some point levelled off, with no additional increase even with categories as large as 32. These findings suggest that two processes are involved in retrieval—a successive scanning mechanism being used with categories small enough to be held in immediate memory (i.e., about 7) and a different mechanism (probably involving little or no successive scanning) operating for larger sets.

How is information retrieved from long-term memory? Although it is too early to give a detailed answer to this question, some recent work does appear to have eliminated some of the logical possibilities. In particular, there is now substantial evidence that the normal process of information retrieval involves little or no successive scanning of items in the memory store. This conclusion is based on a series [illegible] studies which showed that the amount [illegible] time taken to retrieve a word from a large category was either equal to or only slightly greater than the time taken to retrieve a word from a small category. Landauer and Freedman (1968) used a series of nested categories in which each was a subset of the larger one, thus guaranteeing without an actual count that the categories could be ranked according to size (e.g., living thing, animal, bird). They [illegible]und that it took longer to decide that an item was not a member of a large category than to decide that it was not a member of a small category, but there was no appreciable difference in decision times for positive instances. Collins and Quillian (1969), Meyer (1970), and Landauer and Meyer (1972) using somewhat different procedures report that identification took somewhat longer for larger categories. In contrast Neisser (1964) compared identification time for "animals" and "first names" and found no difference in latency. Although most of these studies did find a small difference favoring the small categories, the magnitude of the difference is very small relative to the size of the categories involved. For example, Landauer and Freedman report that the correct identification time with negative instances was 53 msec. longer for large than for small categories. Given the relative size of the sets to be searched, in order for such a small difference to be caused by the additional time it takes to scan the larger categories, successive scanning would have to be enormously rapid, on the order of a thousand words per second, which seems improbable.

More recently Freedman and Loftus (1971) departed from the previous work that had studied identification time and instead measured the speed with which *S* could actually produce a word himself. Instead of giving *S* a stimulus and asking him to decide whether it was a member of a category, he was given a category and asked to produce a word that belonged in it. Several studies by these authors found that the speed of producing a word was independent of the size of the category to which the word belonged. (Freedman & Loftus, 1971; Loftus, Freedman, & Loftus, 1970).

[1] This research was supported by a National Science Foundation grant to Freedman and a National Institute of Health grant to Loftus. The authors are grateful to E. Scheff, J. Nathan, and S. Wiksten for their help in conducting the research, and to T. Nelson for statistical advice and many conversations about item selection.

[2] Requests for reprints should be sent to Jonathan Freedman, Department of Psychology, Columbia University, New York, N. Y. 10027.

Although the results of this previous work are fairly convincing, the use of already existing or natural categories presents certain difficulties. Any existing category is already large or small—it cannot be both. Thus, the size of a category is necessarily confounded with the specific category itself. Accordingly, the effects obtained in these studies could be due to certain characteristics of the specific categories used or to the hierarchical system that exists in the language. Despite the fact that some of these studies have used a large number of different categories and found quite consistent results, it is possible that within the language there are certain characteristics common to all large categories that differentiate them from smaller ones. For example, large categories tend to be higher up in any hierarchical structure since naturally as one goes up the hierarchy a category tends to be more inclusive and therefore have more members. Thus, the lack of effect due to category size could conceivably be due to other factors that tend to be correlated with the size of the category and which for one reason or another facilitate retrieval from large categories.

A number of authors (e.g., Juola, Fischler, Wood, & Atkinson, 1971; Tzeng, 1972) have reported studies in which *S*s learned one or two sets, and were then shown a word and asked if it belonged to the set they had learned. As with other studies of identification time, this work generally found small but significant effects of set size. It does appear, however, that in this research the lists, while learned to criterion, were not learned to the point that the *S*s knew them well. In the Juola et al. study *S*s were given the list 24 hr. before the experimental session and were asked to study it until one successful complete recall could be made. In Tzeng's study, there was a short training session. Thus, although *S*s could recall all of the items on the list, it would be difficult to argue that they had learned them to any great extent.

The level to which *S*s had learned the categories is critical. One can see this, for example, by examining more closely Juola et al.'s two-stage model of recognition. Their first stage involves contact between input and long-term memory necessary to assign a recognition strength to the input. If the item is highly familiar, the *S* is assumed to be able to output a fast response, without checking the item against the memorized list. More specifically, the response to a highly familiar item is assumed to be rapid, error prone, and independent of the number of items in the memorized list. If the presented item is only moderately familiar, however, the *S* is presumed to check the memorized list before responding, and the time to search will be a function of the number of items in that list. The above reasoning appears to be an admission that many of the items were not learned very well. Since we are investigating the difference between short- and long-term memory, and since amount of learning is one element in that difference, it is not clear how useful these results are for our purposes. In any case, it seemed desirable to conduct a study in which *S*s were given extensive training on newly constructed categories and production (as opposed to identification) from these categories was investigated. This paper reports four such experiments.

Experiments I and II

Method

Materials. For Experiment I, 50 words were selected from different categories in the Battig and Montague (1969) norms. Twenty of these words were divided evenly among four lists, so that the 5 words in a list began with each of the letters C, M, P, S, and T. The remaining 30 words (with different initial letters) were randomly assigned as fillers to each list so as to create lists of length: 5 (no fillers added), 10, 15, and 20 words.

For the first *S*, the 50 words were typed on a sheet of paper in such a way that the 5-word list was designated List 1, the 10-word list designated List 2, etc. The next *S* studied lists created by rotating the 30 filler words, so that his List 1 had the same 5 target words as the first *S*'s, but had the 15 filler words which were previously on List 4. Thus, for *S* 2, List 1 had 20, List 2 had 5, List 3, 10 and List 4, 15 words. This process was continued for the next two *S*s so that for each set of four *S*s, each group of 5 target words was in a total list of either 5, 10, 15, or 20 words. For each set of four *S*s, 20 new key words with different initial letters were selected. In all, five such sets of lists were created resulting in a total of 20 unique lists.

The lists for Experiment II were constructed in an identical manner to the lists for Experiment I

with the exception that all words in a particular list came from the same category. The categories were: animal, boy's name, girl's name, or country. Each *S* studied one list from each of these categories. So, for example, for *S* 1, List 1 consisted of five key animals, List 2 of 10 girl's names (5 of them key names), List 3 of 15 boy's names and List 4 of 20 countries.

Subjects. Undergraduates at the Columbia Summer School were paid $6 for 3 1-hour sessions and were promised a bonus if they performed well. In Experiment I, 20 *S*s participated of which only 18 completed the experiment. In Experiment II, 16 *S*s completed all sessions. A $2 bonus was paid to all *S*s who completed the experiment.

Procedure. At the first session the *S*s were told that they would be given four lists of words and that their task was to learn them extremely well. They were supposed to learn what words were in each list and also be able to tell what list a particular word was in. The instructions said, "you should learn these words so that they are almost like second nature." During the first session the *S*s were handed a sheet containing the four lists and given 10 min. to study it. They then were given 5 min. to write down as many of the words as they could remember in the appropriate lists. Next, they were shown cards containing a word followed by the number of its list or the opposite order (e.g., camel—1, 1—camel). The *S*s looked at these cards and said aloud both halves. Each word in the list was presented twice in this manner, once with the word first and once with the number of the list first, a different random order being used for each *S*. Finally, the *S*s once again tried to write down as many words as they could remember in each list. At the end of the session they were given a list and told to study it hard, to practice, and an appointment was made for the next session.

The second session began with a recall task in which the *S* tried to write down as many items in each group as he could remember. (Any *S* who could not recall every item was told to go home, restudy, and a new appointment for the second session was made.) If the *S* recalled perfectly, he then received a number of training tasks in which each of his words was presented to him an equal number of times. In the first task, *E* read each of the words aloud, the *S* telling what group it was in and making up a sentence containing the word. If the *S* gave the wrong group, *E* corrected him. In the next task, *E* read the words aloud again in a different random order and the *S* was required to give another word in the same group. He was told to try to give a different word each time but to respond quickly. Whenever he made an error he was so informed. Finally the recall task was repeated, the *S*s were encouraged to study hard and a third appointment was made approximately two weeks hence. The first two sessions were for purposes of training; the third session produced the essential data. Stimuli were presented on a machine similar to that used in previous studies (cf. Freedman & Loftus, 1971). It consisted of a half-silvered mirror behind which a card was placed. When a light went on behind the mirror the *S* could read what was printed on the card. The left side of the cards could be illuminated at a predetermined interval before the right side. As soon as the right side was lit a timer started which was stopped by the *S* speaking into a unidirectional microphone. In this experiment the *S*s were shown the number of a category to the left and then after a pause of .5 sec. (automatically timed) the right side was lit revealing a single letter. The *S*'s task was to say aloud a member of the indicated category that began with that letter. Subjects were instructed to respond as quickly as possible but to be certain not to make mistakes: "speed is essential but so is accuracy." They were given the following example—"you might see the category 'trees' and the letter 'o' and you should say 'oak' as fast as you can." Subjects were given three practice trials and after each one were asked if they fully understood the instructions. If during these trials they made any errors, they were once again told that they should respond quickly but not make errors. The critical trials consisted of a random presentation of the 20 possible combinations of the four categories and the five critical letters. Thus for each category the *S* gave five responses, one beginning with each of the five selected letters. The *E* did not inform *S*s whether they were correct. If the *S* did not respond within 10 sec., the trial was terminated and the next trial started.

TABLE 1

Reaction Times as a Function of Category Size for Experiments I and II

Experiment	Category size 5	10	15	20
I	1.88	3.98	6.25	5.17
II	1.13	1.94	2.37	1.93

Results

In both experiments all of the *S*s learned the lists well enough to produce the total list without errors during Session 2 and to respond correctly to virtually all of the test trials in Session 2. In addition, on the test trials of Session 3, there were fewer than 4% errors in Experiment I and fewer than 2% in Experiment II. On the other hand, there were a considerable number of trials on which the *S*s gave no response within the 10 sec. allowed. For the 5, 10, 15, and 20 item categories, failures to respond were 4, 12, 29, and 28% respectively for Experiment I, and 2, 2, 12, and 9% for Experiment II. Table 1 presents the mean

reaction times for both experiments as a function of the size of the category. These means are based on median times for each *S* for each category, with the error and no response trials excluded.

The results for both experiments are quite similar. Reaction times are faster for categories of 5 than for larger categories and for Size 10 than for 15, but are actually slower for 15 than they are for Size 20. Using the Newman-Keuls procedure (Kirk, 1969) for simultaneous testing of pairwise comparisons, only the following differences are significant in Experiment I: (*a*) Size 5 is different from all other category sizes, $p < .05$. (*b*) Size 10 and 15 differed, $p < .05$. For Experiment II, size 5 is different from all other category sizes, $p < .05$, while no other differences are significant.

Clearly responses to the smallest category are faster than to any of the others while categories of sizes 10, 15 and 20 do not produce consistent differences in reaction times. If anything, the largest category is somewhat faster than the next largest. This is not a significant effect, but is the opposite of what would be expected from a scanning model. Beyond size 10 there is no evidence that larger categories take longer.

Before discussing the significance of these findings in more detail, it seemed advisable to replicate this study once more. The possibility existed that the results obtained were for some reason dependent on the specific list sizes employed and it therefore seemed advisable to investigate the effects of different size categories. In addition, we wanted to include at least two categories of sufficiently small size so that they would ordinarily be considered to fall within the scope of short-term or immediate memory since it occurred to us that perhaps two different processes were involved in the production of members from the smaller as compared to the larger categories. Accordingly, Experiment III was conducted using lists of size 3, 6, 12, and 24.

Experiment III

Method

Subjects. Twenty-eight *S*s were recruited from a local newspaper advertisement. Most were undergraduates from nearby colleges. They were paid $5 for three sessions and were promised a bonus of $3 if they performed well.

Materials. A set of 12 target words and a se[t of] 33 filler words were selected from the list of nouns provided by Paivio, Yuille, and Madi[gan] (1968) on the basis of three criteria: (*a*) the wo[rds] were four to eight letters in length; (*b*) the tar[get] words began with each of the letters F, S, and [M;] (*c*) none of the filler words began with the let[ters] F, S or M. Lists of length 3 (no filler words add[ed]) 6, 12, and 24 were then constructed accordin[g to] the procedures used for list construction in Exp[eri]ment I.

Procedure. The first session was identical to [that] of Experiment I. After the session *S*s were gi[ven] the lists to take home and study, and an appoi[nt]ment was made for the next session approximate[ly] 1 wk. hence.

The second session began with a recall task [in] which the *S* tried to write down as many items [from] each group as he could remember. Any *S* who d[id] not perform perfectly was sent home and his secon[d] session was rescheduled. During the second sessi[on] a series of 90 stimuli was presented on the machin[e] used in Experiment I. The stimuli consisted of th[e] number of a list (category) and a word. There was an interval of .5 sec. between the category number and word. Subjects were supposed to say "yes" when the word was a member of the category and "no" otherwise. They were asked to respond [as] quickly as possible but to make no errors. Of [the] 90 total trials, half were positive and half w[ere] negative because each word on *S*'s list was paire[d] once with a correct list number and once with an incorrect one. Finally, the recall task was repeated, the *S*s were encouraged to study hard, and a third appointment was made approximately 1 wk. hence.

During the third session, stimuli were pres[ented] on the machine and the *S*s' reaction time for nam[in]g a member of a list beginning with a given lette[r] was measured. The critical trials consisted of a ra[ndom] presentation of the 12 possible combinations of [the] four categories and the three critical letters. T[hus] for each category, the *S* gave 3 responses, o[ne] beginning with each of the 3 selected letters. Reac[]tion times were recorded for all *S*s. The *E* did not inform the *S* whether he was correct.

Results

There were fewer than 4% errors. The percentages of response failures were 2, 8, 20, and 18 for the 3, 6, 12, and 24 item lists, respectively. The mean reaction times for correct responses for production of a list member are presented in Table 2. Using the Newman-Keuls procedure revealed that responses to items in categories of size 3 are significantly faster than responses in all other categories, $p < .05$. No other differences are significant. This

finding parallels that of the first two experiments. With small categories, the larger the set, the slower the reaction times; but at some point, a further increase in size does not produce longer latencies, and there is even a decrease in time for the largest category.

Although the results of these three studies were quite consistent, it was decided to collect data on a wider range of category sizes. Accordingly, Experiment IV involved 6 categories ranging from 4 to 12.

Experiment IV

Method

Subjects. Thirty-six *S*s, mostly college students, were recruited from a local newspaper advertisement. They were paid \$3.50 for 3 sessions, and were promised a bonus of \$3.50 if they performed well.

Materials. A set of 24 target words and a set of 58 filler words were selected from the Paivio, Yuille, and Madigan (1968) nouns on the basis of three criteria: (*a*) the words were from 4 to 8 letters in length; (*b*) the target words began with one of the letters B, F, L, and R and (*c*) none of the filler words began with the letters B, F, L, and R. The 24 target words were divided into 6 sets such that each set had 4 words, each beginning with a different letter. Lists of length 4, 5, 7, 12, 22, and 32 were created using the procedure described in Experiment I. A given *S* received four lists, two short lists and two long lists. Thus, the first *S* received the 4, 5, 12 and 22 word lists, typed on a sheet of paper in such a way that the 4 word list was designated List 1, the 5 word list was designated List 2, and so on. The lists for the next 3 *S*s were created as in previous experiments. For the next group of 4 *S*s, the same process was repeated using the original lists of 4, 5, 12 and 32 words. In all, 36 unique sheets of paper were created, each containing 4 typed lists.

Procedure. Upon arriving at the experimental lab, *S* was told that he would be given four lists of words and that his job was to learn them extremely well. He was given a sheet containing the four lists, told to study it every day and an appointment was made approximately four days hence.

The second session began with a recall task in which *S* tried to write down as many items in each category as he could remember. If he missed more than one word, he was required to go home, relearn the list, and return on a different day. Next *S* was given the standard identification task, using the machine described in Experiment I. He was shown cards containing a word and a list number (e.g., camel—1, 2—shirt) and he responded "yes" into the microphone when the word was a member of the category and "no" when it was not. Negative trials were created by pairing words on the list with incorrect category numbers. The *S* saw an equal number of positive and negative trials, with the word to be categorized printed before the list number on half the trials, and after it on the remaining half.

The third session occurred approximately 2 days later. First *S* was asked to write down as many items from each category as he could remember. Again, if he missed more than one word, his third session was rescheduled. Next, *S* was given the "item-production" task; stimuli were presented on the same machine using the same procedure described previously. In all, a random presentation of the 16 possible combinations of the 4 categories and the 4 key letters was presented to *S*. Thus, for each category, *S* gave 4 responses, one beginning with each of the 4 key letters.

TABLE 2

Reaction Times as a Function of Category Size for Experiments III and IV

Experiment	Category size								
	3	4	5	6	7	12	22	24	32
III	1.44			2.69		3.35		2.91	
IV		2.63	2.71		3.59	3.29	3.95		3.30

Results

There were fewer than 5% errors. The percentages of response failures were 4, 5, 4, 8, 11, and 13 for the 4, 5, 7, 12, 22, and 32 item categories, respectively. The mean reaction times for correct responses for the six category sizes are presented in Table 2. Once again there is an increase in RT's as category size increases from 4 to 5 and from 5 to 7; but at that point no further increase occurs, with 12 and 32 actually being somewhat faster than 7. Using the Newman-Keuls procedure, size 4 and 5 are different from all other category sizes, $p < .05$, but not from each other; no other differences were significant.

General Discussion

At first glance, the present results may appear to be at odds with those reported by Juola et al. (1971). Subjects in their experiment memorized a list of either 10, 18, or 26 words, and then were tested for recognition of those words. The recognition test consisted of presenting a series of single words to *S*, and for each word *S* made a positive or nega-

tive response to indicate whether or not the word was a member of the learned list. Response latency was found to increase as list size increased. However, the Juola et al. paradigm is different from the present one in several major respects. Their *S*s learned a single list of words; ours were tested with a production test. Their *S* could respond on the basis of a subjective judgment of the familiarity of the test item (and this is precisely what Juola et al. claim their *S*s do); while our *S*s could not.

The four experiments reported here have produced consistent results—with small categories, the larger the set, the slower the reaction time; at some point, this effect ceases, and additional increases in set size do not affect latencies. Although the failures to respond increased only slightly with category size for the large categories, one might claim that the stable latencies for correct responses are simply an item selection artifact. The argument is that as category size increases, the strength of any given item in the category progressively decreases, until eventually the item's strength does not allow it to be recalled within the 10 sec. limit. Thus, although the average strength of all items decreases with category size, the weakest items move from being recallable within 10 sec. to being response omissions. This would leave the mean correct-response latency unaffected by category size. However, such an item selection notion would also predict that the entire frequency distribution of correct response latencies would progressively shift in a positive direction as category size increases, and this result did not occur.

A more reasonable interpretation of these results is that two different retrieval processes are involved when items are retrieved from small versus large categories. The smaller categories are within the capacity of short-term memory. The individual is able to keep as many as seven items in his immediate memory store. When he is given the category name and then asked to produce an item beginning with a particular letter, he may put the full category in his immediate store, scan it, and find the appropriate item. This would, of course, produce a category size effect, with larger sets taking longer (Sternberg, 1966). However, once the category is too large for the immediate store, this process is no longer feasible. Instead *S* must rely on a mechanism for retrieving information from long-term memory, and as previous research has indicated, this does not involve a successive search is accordingly not affected by the size o set.

This implies that a process equivalen to that involved in short-term memory ma be activated. The critical element, accordi to our notion, is the size of the set to be sear d. If it is small enough to be held in short- rm memory, it will be; and the search proce ill be the familiar successive scan. If the is too large for the short-term store, a diff nt process will be employed, one involving le or no successive scanning. We would ad at if the normal process for searching long- m memory should fail (i.e., *S* does not find e desired item), it seems plausible that *S* w d again resort to successive scanning and the size of the set would once more be That is, very difficult items from lon memory might force *S* to use some successive processing. Thus, we would pothesize three distinct mechanisms o trieval: (*a*) successive scanning for small s (*b*) nonsuccessive scanning for larger sets long as the items to be found are reas easy or for whatever reason the search cessful within a reasonable period; (*c*) siderably slower and tedious successive for large sets when necessary.

The notion of dual strategies is not n with us; evidence has been found by oth investigators, in different situations. For ample, in an experiment by Mandler and Dea (1969), *S*s were shown a list one word at time. As each additional list item was pr sented, *S* had to recall all of the words pr sented up to that time. With each successi input item, the *S*'s organization of the changed; with categorized lists, in particula these changes were quite dramatic. For th first few trials, an *S* usually began recallin items in the order they were presented. Afte roughly five items had been presented, *S* typically began to organize the items into category clusters rather than recall them in their input order. The typical strategy, then, appears to be as follows: *S*s begin by recalling the specific items on the list in STM; when STM capacity is exceeded, they reorganize the STM information. Rumelhart, Lindsay, and Norman (1972) have incorporated these two strategies into a successful simulation of list learning under these conditions.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replica-

tion and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, 80(3, Pt. 2).

C[illegible] A. M., & QUILLIAN, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240–247.

FREEDMAN, J. L., & LOFTUS, E. F. Retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 107–115.

JUOLA, J. F., FISCHLER, I., WOOD, C. T., & ATKINSON, R. C. Recognition time for information stored in long-term memory. *Perception & Psychophysics*, 1971, **10**, 8–14.

[illegible]RK, R. E. *Experimental design*. Belmont, Calif.: Brooks/Cole, 1969.

[illegible]AUER, T. K., & FREEDMAN, J. L. Information retrieval from long-term memory: Category size and recognition time. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 291–295.

LANDAUER, T. K., & MEYER, D. E. Category size and semantic-memory retrieval. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 539–549.

[illegible]FTUS, E. F., FREEDMAN, J. L., & LOFTUS, G. R. Retrieval of words from subordinate and superordinate categories in semantic hierarchies. *Psychonomic Science*, 1970, **21**, 235–236.

MANDLER, G., & DEAN, P. J. Seriation: The development of serial order in free recall. *Journal of Experimental Psychology*, 1969, **81**, 201–215.

MEYER, D. E. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, **1**, 242–300.

NEISSER, U. Visual search. *Scientific American*, 1964, **210**, 94–102.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Pt. 2).

RUMELHART, D. E., LINDSAY, P. H., & NORMAN, D. A. A process model of long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

TZENG, O. J. L. A triggering system for scanning processes in long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 662–670.

(Received October 26, 1973)

(Continued from page 1084)

[illegible] Poststimulus Study Time on the Recognition of Pictures. George E. Weaver (Department of [illegible]chology, Florida State University, Tallahassee, Florida 32306).

Lat[illegible]d Sound Localization as a Function of Azimuth and Frequency. Frank J. Tolkmitt (Department of Psychology, University of Melbourne, Parkville, Victoria, Australia).

En[illegible]ding Specificity: Semantic Change between Storage and Retrieval Cues. William F. Wallace (De[illegible]ment of Psychology, University of Nevada, Reno, Nevada 89507) and Michael D. Murphy.

[illegible] and Retention of Verbal Lists: Serial Anticipation and Serial Discrimination. Edward A. Wade ([illegible] Little Hall, Department of Psychology, University of Maine at Orono, Orono, Maine) and Michael J. Blier.

Developmental Changes in the Effects of Spacing of Trials in Retardate Discrimination Learning and Memory. Richard D. Sperber (Department of Psychology, University of Connecticut, Storrs, Connecticut 06268).

[illegible]f Input Effects on Subject-Controlled Processes. Carl E. McFarland, Jr. (Department of Psychology, University of Kansas, Lawrence, Kansas 66044) and George Kellas.

Storing and Retrieving Information about Ordered Relationships. George R. Potts (Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755).

Visual Backward Masking by a Flash of Light: A Study of U-Shaped Detection Functions. Alan L. Stewart (New York University School of Education, Deafness Research and Training Center, 80 Washington Square East, New York, New York 10003) and Dean G. Purcell.

Probing Information from Briefly Presented Arrays. Lester A. Lefton (Department of Psychology, University of South Carolina, Columbia, South Carolina 29208).

Regulation and Production of Movement Effects in Exploration-Recognition Performance. David T. [illegible]adrigan (Department of Psychology, Conant Hall, University of New Hampshire, Durham, New Hampshire 03824) and G. Alfred Forsyth.

Context Effects in Sentence Memory. Marcia K. Johnson (Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11790), Theodore J. Doll, John D. Bransford, and Robert Lapinski.

Visual Detection and Visual Imagery. M. J. Peterson (Department of Psychology, Psychology Building, Indiana University, Bloomington, Indiana 47401) and S. E. Graham.

Free-Recall Performance as a Function of Overt Rehearsal Frequency. William F. Battig (Institute for the Study of Intellectual Behavior, Muenzinger Building, University of Colorado, Boulder, Colorado 80302), Gerald O. Einstein, James W. Pellegrino, and Michele S. Mondani.

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1092–1099

STIMULUS AND RESPONSE FREQUENCY AND SEQUENTIAL EFFECTS IN MEMORY SCANNING REACTION TIMES [1]

JOHN THEIOS [2] AND DENNIS G. WALTER

University of Wisconsin—Madison

Target set size, stimulus frequency, and response–stimulus interval were varied in a design first used by Sternberg. Significant effects on RT were found due to stimulus frequency, stimulus sequence, response frequency, response sequence, and the interaction between stimulus frequency and target set size. Response–stimulus interval affected only the RT intercept. The data are difficult to account for assuming a serial exhaustive memory scanning process, but can easily be accounted for using a self-terminating scanning process.

Theios, Smith, Haviland, Traupmann, and Moy (1973) have demonstrated large stimulus presentation frequency effects on reaction time (RT) in a Sternberg (1967) type fixed-set memory scanning task. Mean RT to both positive and negative stimuli increased approximately 85 msec. as stimulus presentation probability was decreased from .30 to .05. Theios et al. interpreted these results as supporting a serial self-terminating memory scanning process such as the model proposed by Theios (1973a, 1973b). It is assumed that memory representations of both target and nontarget stimuli are organized probabilistically, trial-by-trial into a limited capacity, push-down memory stack on the basis of stimulus occurrence. More frequent stimuli (both target and nontarget) result in relatively fast RTs since they are more likely to have occurred recently and have their S–R memory pairs located relatively early in the serial scan. Reaction time to infrequent stimuli will be relatively slow since their S–R memory pairs will be located relatively late in the serial scan.

In addition, the Theios (1973b) self-terminating memory model predicts that there should be both stimulus and response sequential effects in memory scanning RTs similar to those found by Bertelson (1965), Rabbitt (1968), Remington (1969), and Smith, Chase, and Smith (1973) in choice RT. Reaction time to a stimulus in a memory scanning task should decrease as a function of recency and frequency of either stimulus or response occurrence in any random sequence of trials. A response frequency as well as a stimulus frequency effect on RT follows directly from the sequential properties of the process.

Finally, the Theios (1973b) model predicts an interaction in RT between stimulus frequency and target set size, since both these variables are assumed to affect the memory scanning process.

The exhaustive scanning process suggested by Sternberg (1966, 1967) can account for response frequency and response sequential effects by a response process such as Sternberg's (1969) binary decision or response translation stages. However, the exhaustive scanning model does not have any explicit mechanism to account for stimulus frequency or stimulus sequential effects in RT. The exhaustive scanning model might possibly account for stimulus frequency and sequential effects by assuming that stimulus recency affects the stimu-

[1] This report is based upon a research thesis directed by the first author and conducted by the second author in partial fulfillment of the requirements for the degree of Master of Science.

The research was supported in part by U. S. Public Health Service Grant MH-19006 from the National Institute of Mental Health. The second author was supported by a University of Wisconsin Fellowship. Use of the Madison Academic Computing Center was made possible in part through grants from the Wisconsin Alumni Research Foundation and the National Science Foundation.

We would like to express our gratitude to Melvyn C. Moy and Peter G. Smith for assistance in computer programming and data analysis.

[2] Requests for reprints should be sent to John Theios, Department of Psychology, University of Wisconsin—Madison, Psychology Building, Charter at Johnson, Madison, Wisconsin 53706.

lus encoding process such that more frequent and recent stimuli are encoded faster than less frequent and less recent stimuli. This would predict stimulus frequency effects on RT to target and nontarget stimuli, stimulus sequential effects, and *no* interaction between stimulus frequency and target set size. This follows since stimulus frequency would affect the encoding process and target set size would affect the memory scanning process, and the effects of the two variables would add rather than interact (cf. Sternberg, 1969).

The present experiment and analyses were conducted to provide further evidence on the question of whether memory scanning is self-terminating or exhaustive. A critical question is whether there will (self-terminating prediction) or will not (exhaustive prediction) be an interaction in RT between stimulus frequency and target set size. The design of the experiment was effectively a replication on our equipment of the essentials of the Sternberg (1967) intact stimulus condition. In addition, however, we analyzed the RT data for stimulus and response probability effects and sequential effects and for an interaction involving stimulus probability and target set size.

Method

Subjects. The *Ss* were 48 University of Wisconsin undergraduates (30 females and 18 males) who participated in order to fulfill an introductory psychology course requirement. It was required that all *Ss* be right-handed.

Apparatus. The apparatus was the same as that described by Theios et al. (1973). The stimuli were the set of digits 0 through 9 presented on an Industrial Electronic Engineers, Inc., Series 10 rear-projection visual readout unit. The *S* sat in an arm chair with a response button mounted in each arm beneath the index finger. A Digital Equipment Corporation PDP-8 computer was programmed to select at random the stimulus for each trial with the probability for the selection of any particular stimulus on a given trial preset by *E*. The computer also measured the RT and timed the response–stimulus interval.

Design. The 48 *Ss* were divided into two groups of 24 each. One of the groups was instructed to press the right-hand button whenever a positive set stimulus occurred and the left-hand button otherwise. The other group pressed the left button on positive trials and the right one on negative trials.

The 10 stimuli were partitioned into positive and negative sets in the same manner as in the Sternberg (1967) study. For a given *S*, 1 digit was presented on approximately 4/15 of the trials, 2 others were presented on approximately 2/15 of the trials each, and the remaining 7 each occurred on approximately 1/15 of the trials. Each *S* was run through four conditions with positive memory set size (M) equal to one, two, three, or four. Each of the 10 digits was assigned to the positive set for one condition and to the negative set for the three others. For M = 1, 2, and 4 the stimuli were assigned to the positive set so that the probability of a positive response occurring was 4/15. The M = 3 condition was always presented first, as a practice condition, and here the probability of a positive response was 3/15 (the digits assigned to the positive set in this condition were the three 1/15 probability digits not used in the M = 4 positive set). Only data from the conditions where M = 1, 2, or 4 were analyzed. Each of the six possible orders of presentation of the M = 1, 2, or 4 conditions was assigned to four *Ss* within each group. Two blocks of trials were run for each condition, with 108 trials in each block. For one block, the time between *S*'s response to one stimulus and the onset of the next stimulus (the response–stimulus interval or RSI) was .5 sec. For the other block, the RSI was 2 sec. Each of the two presentation orders of these blocks was assigned to two members of the four-subject subgroups. The actual assignment of particular digits to the positive set in each condition followed the example of Sternberg (1967).

Procedure. The instructions urged *S* to respond as fast as possible without making errors. The number of permissible errors was never made explicit. The *S* was told that if he made too many errors in any one block of trials, he would be given additional blocks of trials until his error rate was acceptably low. Actually, the largest number of errors permitted in a block of 108 trials was nine. The *S* was then told the numbers in the positive set. After a block of trials *S* was allowed to relax while the error data was being typed out on the teletype. If the number of errors was within the permissible limit, the next block of trials, with the same positive set but with the other level of RSI, was begun. If the number of errors was too large, the trial block was repeated. This procedure was continued, alternating short and long RSI, through all four memory set size conditions until all eight different trial blocks were run.

Results and Discussion

The overall error proportion was less than .03, and trials on which errors occurred were not included in the following analyses. The effect of the between-subjects variable (dominant vs. nondominant hand on positive set) was examined by an analysis of variance along with several of

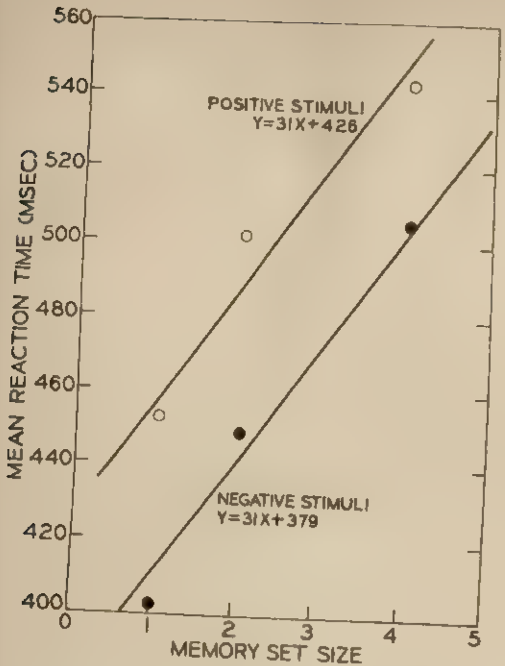


FIGURE 1. Mean reaction time for positive (target) and negative (nontarget) stimuli as a function of memory set size with the best-fitting pair of parallel linear functions.

the other variables. This variable was found to be nonsignificant ($F < 1.00$), and in all further analyses all *S*s were treated as one group.

Figure 1 shows the increase in mean RT over memory set size for positive and for negative trials plotted against the best-fitting pair of parallel straight lines. Separate analyses of variance were performed for the positive and negative data to determine whether the null hypothesis of linearity could be rejected for either. It was found that both the positive and negative data points differed significantly from their respective best-fitting straight lines. For positive data points, $F(1, 94) = 12.88$, $p < .001$, and for negative data points, $F(1, 94) = 6.08$, $p < .05$. However, it was possible to account for 97% of the variance among the six means by the hypothesis of parallel linear functions. The data are of the same general form as Sternberg's (1967) data, and the present study used four times as many *S*s as Sternberg's, meaning that for the present study there is a great deal more power for the analysis of va[r]iance test of the linearity hypothesis.

Stimulus Frequency Effects

An important test of the theore[tic]al models is whether or not stimulus pre[sen]tation probability effects can be found [be]tween items for a given memory set [size]. Figure 2 shows the mean RT for nega[tive] items at each of the levels of probabi[lit]y for each target set size. Note that R[T] is higher for low probability items under e[ach] set size condition. Two analyses of [vari]ance were performed on the data re[pre]sented graphically in Figure 2, the firs[t on] mean RT for negative items with pr[oba]bility 1/15 and 2/15 and with 4 and 1 it[ems] in the positive set and the second on m[ean] RT for negative items with probabil[ity] 4/15 and 1/15 and set size 2 and 4. T[he] pattern of significant effects was the sam[e] for both analyses. With 1 and 47 degree[s] of freedom in each comparison, there were significant main effects due to target set size, $F = 234.28$ and $F = 70.94$, $ps < .001$ for the first and second analysis, respectively, stimulus probability, $F = 8.29$, $p < .01$ and $F = 27.26$, $p < .001$, response stimulus interval $F = 28.24$ and $F = 39.09$, $ps < .001$, along with only one significant interaction, that involving set size and probability, $F = 5.22$, $p < .05$ and F [=] 8.24, $p < .01$.

The finding of a probability effect among the negative items could not be predicted on the basis of an exhaustive scan of the positive set as proposed by Sternberg (1967). However, the finding is consistent with a model in which both positive and negative items are represented in memory and mean RT is inversely related to stimulus frequency and recency (Theios et al., 1973). On the other hand, the possibility must be considered that the probability effect does not reflect the nature of the serial comparison stage, but rather one of Sternberg's (1969) other processing stages, presumably the encoding stage. However, the observed interaction between stimulus probability and target set size leads to the conclusion, on the basis of the Sternberg analysis, that set size and stimulus probability affect the same processing stage. In

the Sternberg papers, the assertion that the serial comparison stage is exhaustive is based on the parallel increase in mean RT for positive and negative items as target set size is increased. The target set size variable and the serial comparison stage are strongly linked. Therefore, the finding of the interaction between stimulus probability and target set size is not compatible with the assumption of a serial comparison stage unaffected by stimulus probability.

One way in which an exhaustive scanning model might be used to explain the probability effects found in the present study is that the serial comparison stage could be presumed to consist of an exhaustive scan of the target set but with the scanning rate faster for high probability probe stimuli. However, the basis for Sternberg's assertion of an exhaustive scan rather than a self-terminating scan is the parallel increase in mean reaction time for positive and negative items over increases in memory set size, rather than an increase which is twice as great for negative items as for positive items. But that result was obtained in a situation (the present design) where target set size and relative stimulus probability were confounded. If stimulus probability affected scan time significantly, it would mean that the removal of the probability effect (so that the "pure" target set size effect could be seen) would result in a reduced slope for mean RT to positive items since items which are in small target sets are also highly probable in the Sternberg (1967) design. Also, there would be an increased slope for mean RT to negative items since the high probability stimuli are negative under the larger set size conditions. The consequence, therefore, of adopting this particular modification of the exhaustive scanning model is to destroy the original rationale for proposing the model. Since stimulus probability affects mean RT, the rationale for the Sternberg (1967) exhaustive scanning model is called into question regardless of whether stimulus probability is presumed to affect the scanning stage, the encoding stage, or any other stage, since the parallel linearity of his obtained functions must be assumed to be

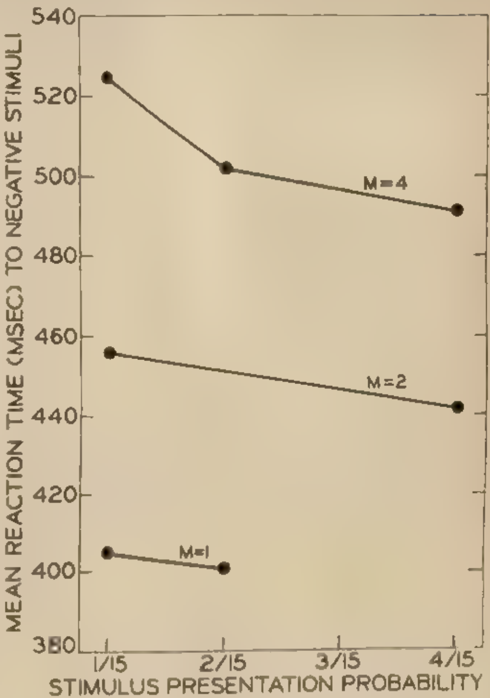


FIGURE 2. Mean reaction time to negative (nontarget) stimuli as a function of stimulus presentation probability and target set size (M).

partially due to variations in stimulus probability as well as to target set size.

Response–Stimulus Interval

The mean RT for the long (2.0 sec) response–stimulus interval (RSI) was significantly higher than for the short (.5 sec.) RSI. The fact that RSI does not have a significant interaction with either target set size or stimulus probability implies that RSI does not affect serial comparison, but rather some other processing stage. It is possible that variation in RSI affects *S*'s "readiness" in a continuous information-processing task. With RSI at .5 sec. the task demands complete attention and continuous concentration for a period of less than 2 min. for a block of 108 trials. For an RSI of 2.0 sec., 108 trials take nearly 5 min. and the length of time between trials is enough so that the *S*'s attention may move away from the task until it is brought back by the onset of the next stimulus. It is not necessarily the case that the RSI affects a precomparison stage.

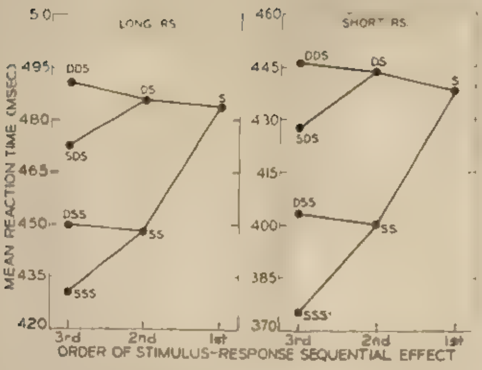


FIGURE 3. Mean reaction time (msec.) as a function of sequence of stimuli for long and short response–stimulus intervals, respectively. [Let S represent any particular stimulus and D represent any stimulus different from S. The point S represents the mean reaction time to stimulus S over all trials. The points DS and SS represent the mean reaction time (over all trials and all stimuli) to stimulus S given that a different (D) or the same (S) stimulus was presented on the previous trial. Likewise, the letter triples represent the mean reaction time (over all stimuli and trials) to stimulus S given that a different (D) or the same (S) stimulus was presented on the previous two trials. For example, the point DDS represents the mean reaction time to a stimulus S given that another stimulus (or two other stimuli) were presented on the preceding two trials.]

It may instead affect response selection or evocation time.

Sequential Effects

If the ordering of a self-terminating memory stack is based on frequency and recency, it follows that the same stimulus occurring twice or more in succession would increase the probability of that stimulus representation moving to an earlier position in the memory buffer. Accordingly it is useful to determine if mean RT is affected by whether the stimulus on a given trial is the same or different from the stimuli on immediately preceding trials. Figure 3 shows the stimulus–response sequential effects in the RT data. The figure shows the mean RT to a stimulus (S) as a function of the stimuli which occurred on immediately preceding trials. If the same (S) stimulus is repeated, reaction time systematically decreases. On the other hand, if the stimulus S has been preceded by different (D) stimuli the rea[ction] time systematically increases.

Response sequential effect. The sequ[en]tial effects in Figure 3 could be due t[o] either stimulus repetitions or response repetitions. That there are response sequential effects in the data is shown in Figure 4. Here RT to a stimulus (S) is given as a function of whether the same (S) response or a different (D) response was given on immediately preceding trials. H[owever,] these response sequential effects are [...] confounded, to some degree, with stimulus repetitions.

"Pure" stimulus sequential effect. [In] order to obtain measures of the pure sti[mu]lus and response sequential effects, [we] considered only runs of trials in which [the] same response is given repeatedly. Fig[ure] 5 shows the pure stimulus and respo[nse] sequential effects. The top branch sho[ws] the pure response repetition effect whe[n] the stimulus is not repeated, but the response is repeated. Reaction time s[...]

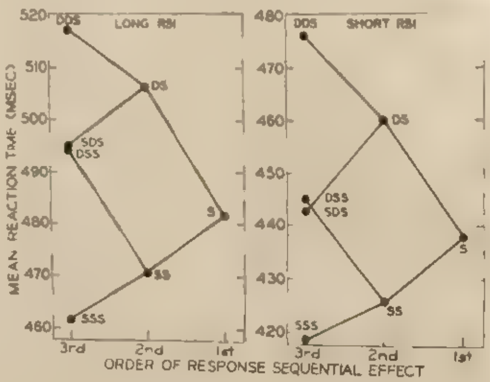


FIGURE 4. Mean reaction time (msec.) as a function of sequence of responses for long and short response–stimulus intervals, respectively. [Let S represent any particular *response* and D any response different from S. The point S represents the mean reaction time for response S over all trials. The points DS and SS represent the mean reaction time (over all trials and responses) for response S given that a different (D) or same (S) response occurred on the previous trial. Likewise, the letter triples represent the mean reaction time (over all trials and responses) for response S given that the same (S) or different (D) responses occurred on the previous two trials. For example, the point DSS represents the mean reaction time for response S given that the same (S) response was made on the previous trial but a different (D) response was made two trials previously.]

tematically decreases with response repetition. The lower branch shows the decrease in RT when both stimulus and response are repeated. As can be seen, the decrease in RT is much greater for stimulus *and* response repetition than for response repetition alone. The difference between the nodes of any two adjacent branches of the sequential tree diagram is a measure of the pure stimulus repetition effect. On the basis of these analyses, it is obvious that there are both stimulus and response sequential effects in memory scanning data, at both long and short response–stimulus intervals.

Application of a Self-Terminating Model

Theios et al. (1973) discussed three models of a self-terminating memory scanning system. The first model had an unlimited capacity serial scanning buffer, the second model had a fixed sized, limited capacity scanning buffer, and the third model had a variable sized, limited capacity scanning buffer equal to the target set size. Each of these three models had two versions, one in which movement up in the memory buffer was the same for target and nontarget stimulus representations ($a_+ = a_-$) and one in which target (positive) representations moved up faster than nontarget (negative) representations ($a_+ > a_-$). The assumption of the equality of the movement parameter can be rejected on the basis of the data from the present experiment. Assuming a_+ equal to a_-, it was not possible to get enough slope in the predictions for positive RT as a function of target set size. Thus, with a_+ greater than a_-, target representations on the average will be located earlier in the memory buffer than representations of nontarget stimuli of equal frequency.

The fact that there is a target set size effect in the data of Figure 2 is sufficient to reject the assumptions of either an unlimited capacity buffer or a fixed sized, limited capacity buffer. The data suggest that the buffer capacity increases with target set size.

Taken together, the data of the present experiment suggest the variable sized memory stack model proposed by Theios et al.

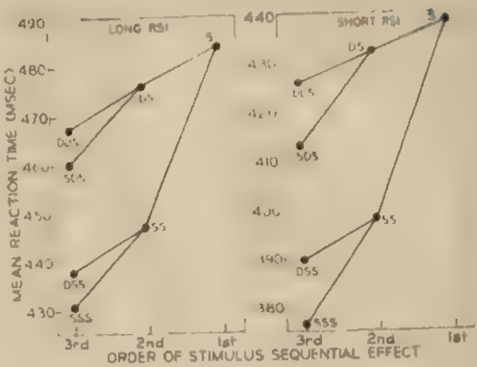


FIGURE 5. Mean reaction time (msec.) as a function of sequence of stimuli (for long and short response-stimulus intervals, respectively) during runs of the same response. [Let S represent any particular stimulus and D represent any stimulus different from S. The point S represents the mean reaction time to stimulus S over all trials. The points DS and SS represent the mean reaction time (over all trials and all stimuli) to the stimulus S given that a different (D) or the same (S) stimulus was presented on the previous trial and given that the response on both trials was the same. Likewise, the letter triples represent the mean reaction time (over all stimuli and trials) to stimulus S given that a different (D) or the same (S) stimulus was presented on the previous two trials and given that the response on all three trials was the same. For example, the point DDS is the mean reaction time to a stimulus S given that different stimuli requiring the same response were presented on the two previous trials.]

(1973). Using a Monte Carlo computer-simulation program, separate quantitative predictions were obtained from the self-terminating memory scanning model described in Theios (1973b) and Theios et al. (1973) for the long and short response stimulus interval conditions. A statistical response protocol corresponding to the self-terminating process was obtained for every response protocol given by a real *S* in the present experiment, under the assumption that the *S*s were homogeneous with respect to the values of the parameters of the process. The size of the memory stack ($Z - 1$) was set equal to the target set size and the time to retrieve a response from long-term memory given an unsuccessful exhaustive scan of the buffer (t_Z) was set equal to the buffer comparison time (t). Chandler's (1969) subroutine STEPIT was used to estimate the values of the remaining five parameters. They were the RT

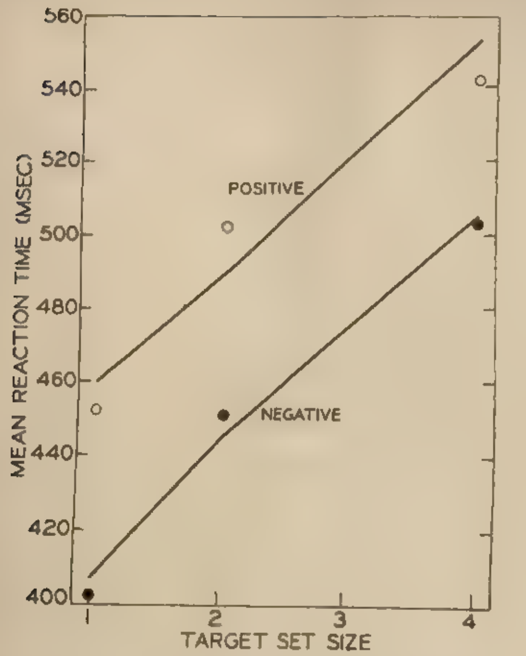


FIGURE 6. Mean reaction time as a function of target set size for positive (target) stimuli (open circles) and negative (nontarget) stimuli (closed circles). [The lines are predictions obtained from a Monte Carlo computer simulation of the self-terminating memory scanning model of Theios, 1973b.]

intercept (the combined encoding and negative response time, C), the difference between positive and negative response selection and output times (D), the memory comparison time (t), the probability of a positive stimulus–response representation moving up to the top of the memory stack after that stimulus is presented (a_+), and the probability of a negative stimulus–response representation moving up to the top of the stack after that stimulus is presented (a_-).

Table 1 shows the mean RT values from :he computer simulation as well as the nean RTs from the present experiment. \ goodness-of-fit analysis of variance was erformed to determine whether the ob- erved mean RTs differed significantly from hose obtained from the two computer mulations. There are 20 comparisons nean differences) involved, and a total of) parameters were estimated from the ata to obtain the predicted means. Thus, ie numerator degrees of freedom are 10. The error term for the F ratio is the average of the squares of the standard errors of the observed means, that is, the within-cell variance, which has 940 degrees of freedom. The test indicated no significant differences between the simulated means and the obtained means, $F(10, 940) = 1.19$, $p > .05$. It is reassuring to note the parameter invariance obtained from the two independent simulations. The parameter estimates for the long and short RSI conditions differ appreciably only in the RT intercept (C). The fit of the model would be almost as good if the long and short RSI conditions were simulated together, using common values for the remaining parameters, i.e., if only 6 parameters would have been estimated.

Figure 6 shows that, contrary to common belief, the self-terminating process predicts that mean RT is an essentially linear function of target set size with approximately equal slopes for target and nontarget stimuli. Figure 7 shows that the self-terminating process predicts the obtained interaction in RT between target set size and stimulus frequency.

Inspection of the estimated values of the parameters in Table 1 indicates that the scanning rates are typical (42 msec.) and

TABLE 1

OBSERVED MEAN REACTION TIMES (MSEC.), STANDARD ERRORS, MEANS FROM TWO SIMULATIONS OF THE SELF-TERMINATING MODEL, AND PARAMETER VALUES USED IN THE SIMULATIONS

Condition	Short response-stimulus interval			Long response-stimulus interval		
	Simulated	Observed	SE	Simulated	Observed	SE
Target Set Size 1						
Positive 4/15	435	434	8	485	471	10
Negative 2/15	384	376	8	429	426	12
Negative 1/15	386	380	8	431	429	11
Target Set Size 2						
Positive 2/15	468	480	12	512	524	14
Negative 4/15	415	414	10	459	470	12
Negative 1/15	427	439	10	470	482	13
Target Set Size 4						
Positive 1/15	533	529	10	573	560	11
Negative 4/15	464	472	11	508	510	11
Negative 2/15	489	488	12	530	516	11
Negative 1/15	504	505	12	544	544	13
Parameter values						
C		300			349	
D		79			79	
t		43			41	
a_+		.34			.43	
a_-		.06			.08	

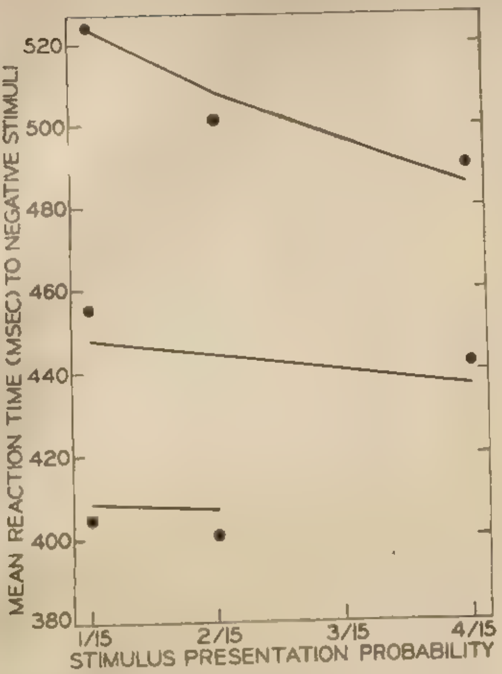


FIGURE 7. Mean reaction time (dots) to nontarget stimuli as a function of stimulus presentation probability and target set size. [The lower two points are for a target size of one, the middle two points are for a target size of two, and the upper three points are for a target set size of four. The lines represent predictions obtained from a Monte Carlo computer simulation of the self-terminating memory scanning model of Theios (1973b). The data clearly show the significant obtained and predicted interaction between stimulus probability and memory set size.]

that the a_+ estimates are larger than the a_- estimates reflecting the greater emphasis on the target items in the scanning process. The 79-msec. difference between positive and negative RT intercepts reflects the fact that negative responses were required on 11/15 of the trials whereas positive responses were required on only 4/15 of the trials. Sternberg (1969) has reported similar results. The 79-msec. difference can be accounted for by bias in what Theios (1973b) has called the response program selection process.

CONCLUSIONS

The results clearly indicate that in the Sternberg (1967) fixed-set memory scanning task there are significant (*a*) stimulus frequency effects, (*b*) stimulus sequence effects, (*c*) response frequency effects, (*d*) response sequence effects, (*e*) target set size effects, and (*f*) effects due to an interaction between stimulus frequency and target set size. The obtained interaction is especially difficult to account for using an exhaustive scanning process. However, the observed pattern of results is easy to account for using a self-terminating process (Theios, 1973b; Theios et al., 1973) as was shown by a computer simulation of the present experiment. When viewed against the background of the research literature in the choice reaction time and memory scanning areas, the present results suggest the general theoretical process of human perception, cognition, decision making, and response selection proposed by Theios (1973a, 1973b).

REFERENCES

BERTELSON, P. Serial choice reaction-time as a function of response versus signal-and-response repetition. *Nature*, 1965, **206**, 217–218.

CHANDLER, J. P. STEPIT: Finds local minima of a smooth function of several parameters. (CPA 312) *Behavioral Science*, 1969, **14**, 81–82.

RABBITT, P. M. A. Repetition effects and signal classification strategies in serial choice-response tasks. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 232–240.

REMINGTON, R. J. Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology*, 1969, **82**, 250–275

SMITH, E. E., CHASE, W. G., & SMITH, P. G. Stimulus and response repetition effects in retrieval from short-term memory: Trace decay and memory search. *Journal of Experimental Psychology*, 1973, **98**, 413–422.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

STERNBERG, S. Two operations in character recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53.

STERNBERG, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, 1969, **30**, 276–315.

THEIOS, J. The components of response latency in human information processing. Paper presented at the Fifth International Symposium on Attention and Performance, Stockholm, Sweden, July 1973. (a)

THEIOS, J. Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 7. New York: Academic Press, 1973. (b)

THEIOS, J., SMITH, P. G., HAVILAND, S. E., TRAUPMANN, J., & MOY, M. C. Memory scanning as a serial, self-terminating process. *Journal of Experimental Psychology*, 1973, **97**, 323–336.

(Received October 26, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1100–1107

STIMULUS PROBABILITY AND STIMULUS SET SIZE IN MEMORY SCANNING[1]

IRVING BIEDERMAN[2] AND E. WEBB STACY, JR.

State University of New York at Buffalo

In many recent studies of speeded scanning of immediate memory, variations in the size of the positive set (s) were confounded with variations in the probability (P) of the individual items within the positive set: As s increased, P decreased. The present experiment sought to determine whether the effect on RT attributed to s could be accounted for by variations in P. This was accomplished by factorially varying both s and P. Probability effects were confined to items in the positive set and were insufficient to account for the effect of s. The results are discussed in terms of a model in which s and P affect different information-processing stages. The s affects the number of comparisons between the encoded item and the items stored in the memory of the positive set, as proposed by Sternberg. The P affects response selection—information as to the particular digit that was presented is available to the mechanisms for response selection along with the knowledge that there was or was not a match. The response selection mechanisms are assumed to be biased in tune with the P values of the items within the positive set.

The number of things that one has to think about and the expectancy as to the likelihood of occurrence of these things—stimulus number and stimulus probability—have long been regarded as fundamental variables in the study of cognition. The common finding that longer RTs would be produced by an increase in the number of possible stimuli or a decrease in stimulus probability was a result that was compatible with most theories of stimulus recognition. Discriminating among the various theoretical accounts for these effects has been a more elusive task.

One class of models holds that variations in stimulus probability and stimulus number affect only a single commodity such as information (in bits) or repetitions. Examples of such models are those that posit a single decision maker which operates on stimulus information in bits (e.g., Briggs & Johnsen, 1973; Briggs & Swanson, 1970; Hick, 1952) or sequential position in a stack (Theios, Smith, Haviland, Traupmann, & Moy, 1973).

In contrast, Sternberg (1966, 1967) has advocated a model in which the time to scan immediate memory is affected by the number of possible items in memory—not by their probability of occurrence. In a series of character classification experiments, Sternberg (1966, 1967) found that the time to classify a character (e.g., a digit) as a member of an arbitrary subset of characters (the memory or positive set) was a linear increasing function of the number of items, s, in that subset. Moreover, the slopes for the positive (match) and negative (nonmatch) functions were identical. From these data, Sternberg postulated a high speed (25 items/sec), exhaustive, sequential scan of the items in the positive set. However, in Sternberg's experiments, the number of items in the positive (memory) set s was confounded with the probability (P) of the individual items in that set: As s increased, P decreased. By interpreting his results solely in terms of stimulus number, Sternberg was implicitly assuming that there was either

[1] This research was supported by a grant from the Research Foundation of State University of New York and a National Institute of Mental Health Special Fellowship to the senior author. Appreciation is expressed to Saul Sternberg, Edward E. Smith, and Roberta L. Klatzky for their many helpful comments, and to Dennis D. Howard, John J. Cleary, and Myra P. Leffel for their aid in the data analysis.

[2] Requests for reprints should be sent to Irving Biederman, Department of Psychology, State University of New York at Buffalo, 4230 Ridge Lea Road, Buffalo, New York 14226.

no effect of P[3] or else that P affected some tage other than memory scanning. This onfounding, which Sternberg explicitly acknowledged, was a necessary consequence of controlling for stimulus and response entropy and response probability. For example, in Sternberg's (1967) experiment, when $s = 1$, 2, and 4, the individual items had P-values of 4/15, 2/15, and 1/15, re- ctively, so that the probability of a Yes sponse was always 4/15. Because of the founding between s and P, Biederman l Zachary (1970) suggested that the ower RTs to the larger set sizes might, in part, represent adjustments to stimulus probability within the positive set. (Sternberg, 1966, found no effect of increasing number of negative set items from 2 15, with negative response probability ld constant, so if probability adjustments re at play, they would have to be to sitive set stimuli.) An effect of P had dy been documented in other experiments (e.g., LaBerge & Tweedy, 1964), in faster RTs were produced by more able stimuli even when the different uli were assigned to the same response. he present experiment was designed to al whether such probability adjustments would be found in the Sternberg memory-scanning task. Specifically, the probability of the items *within* a given sitive set was varied. If it is assumed that P affects memorial comparisons and the memorial comparisons proceed at a fixed rate, then an exhaustive, sequential nning theory would predict that such a bability imbalance would have no effect on RTs.

Recently, however, Klatzky and Smith (1972) argued that an exhaustive scanning theory could accommodate such a result by assuming that the adjustments to probability were made at some stage other than the comparison stage; viz., a stimulus encoding stage. The exhaustive scanning theory and the additive factors method (Sternberg, 1969) then predict that variations in P and s should produce additive effects on RT. Such a result was reported by Klatzky and Smith in a letter classification task. Also compatible with an encoding effect of P is Miller and Pachella's (1973) finding that P and stimulus degradation (which would be expected to affect encoding) interact. The RTs to items of low probability of occurrence were more adversely affected by the presence of visual noise than high-P items. Since Sternberg (1967) found that degradation did not interact with variations in positive set size (i.e., memory scanning), the Klatzky and Smith (1972), Miller and Pachella (1973), and Sternberg (1967) experiments are all compatible with a two-stage theory, encoding and memory scanning, for the effects of P and s, respectively.

The design of the present experiment compared the effect of P on positive RTs with the effect on negative RTs. An encoding theory of P-effects would predict that response type would not interact with P. Theios et al. (1973), Krueger (1970), and Miller and Pachella (1973), in fact, reported such a result. However, these investigators all employed positive and negative sets that were of equal size. Such a procedure might well lead to S's functional (i.e., positive) set not being the one specified by E. When sets are of equal size, there is little utility for S to confine his scanning to the set which E designated as positive. Under such conditions, S might reasonably be expected to sometimes scan one and sometimes scan the other set. Or else, S could scan an amalgamation of sets (Hawkins & Hosking, 1969) by combining the high-P items from the negative set with the positive set. The design of the present experiment employed a positive set that was smaller than the negative set hence, less scanning would be required if S confined his scanning to the positive set.

Method

Subjects. Forty-eight undergraduate students at the State University of New York at Buffalo participated in the experiment as part of their course requirement in introductory psychology.

Stimuli and apparatus. The stimuli were the 10 digits, 0, 1, . . . , 9, generated by an IEEE in-line display (Model 10-0K21-1820-L) set in a 3×3 ft. gray frame. The background, pre- and postexpo-

[3] As defined here, P is the probability of occurrence of a particular stimulus.

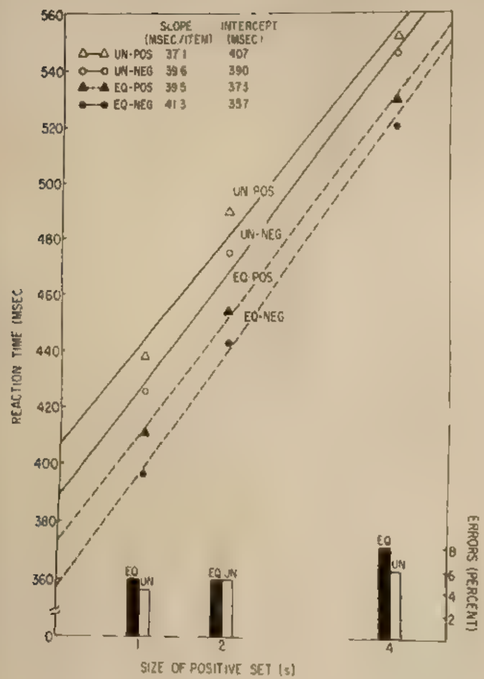


Figure 1. Mean correct reaction time as a function of group (EQ vs. UN), response (POS vs. NEG), and set size. (Slopes and intercept constants shown on figure represent best least squares fit. Lower bars represent error rates as shown on scale on the right of the figure.)

sure fields were black. The *S*'s index fingers rested on plastic keys which, when depressed, activated a microswitch which stopped a timer that was started by the onset of the digit. A warning buzzer of .5-sec. duration sounded 1 sec. before each stimulus presentation of .5 sec. The interstimulus interval was 5 sec. After each response, *E* verbally provided error and speed feedback (to the nearest .01 sec.). Approximately 2 min. intervened between blocks of trials.

Design and procedure. Each *S* was fully instructed about the nature of the task and the relative frequencies of each of the 10 digits. A bar graph display of these relative frequencies remained in *S*'s view throughout the experiment. Left and right keys were assigned to Yes and No responses according to *S*'s preference.

The design closely followed that of the first session in Sternberg's (1967) experiment. The first block for all *S*s always consisted of 45 trials with a positive memory-set size of three digits. The second through fourth blocks each consisted of 90 trials with memory set sizes of one, two, or four digits in counterbalanced order. Half of the *S*s were run in a condition (UN) where, for memory sets of Size 2 and 4, one of the digits in that set was more probable than the other digit(s). The remaining *S*s, like those in Sternberg's experiment, had memory sets where the positive digits were equiprobable (EQ). The stimulus sets and probabilities are shown in Table 1. For Set Size 2 in the EQ condition, both digits had a probability of 2/15. In the UN condition one of the digits had a probability of 3/15 and the other digit had a probability of 1/15. For Set Size 4 in the EQ condition, each of the four digits had probabilities of 1/15. In the UN condition, one of the four digits had a probability of 2/15 and the remaining three digits each had a probability [illegible] For Set Sizes 1, 2, and 4 the probability of a [illegible] response was 4/15. For $s = 3$, all *S*s had a [illegible] probable condition where the probability of [illegible]tive response was 3/15. (Therefore, the probability for any one digit when $s = 3$ was 1/15.) For EQ *S*s, therefore, 1 of the 10 digits had a probability of 4/15, 2 had probabilities of 2/15, and the remaining 7 had probabilities of 1/15. In the UN condition, 1 digit had a probability of 4/15, 1 had a probability of 3/15, 1 had a probability of 2/15, 4 had probabilities of 1/15, and 3 had probabilities of 2/45.

Each *S*'s positive sets were selected from one of the three groupings used in the Sternberg (1967) experiment. The high and low *P*s were balanced over the different digits. Thus, for example, one-third of the *S*s had a grouping that when $s = 1$ the positive digit was "5"; when $s = 2$, the positive digits were 4 and 9; and when $s = 4$, the positive digits were 0, 1, 3, and 7. In the UN condition, each digit in $s = 2$ and 4 occurred as the more probable digit an equal number of times. For example, for eight of the *S*s in the UN condition, when $s = 2$ the more probable ($P = 3/15$) digit for four of these *S*s was "4" and for the remaining four *S*s the more probable digit was "9." The assignment of *P*s to digits remained constant throughout *S*'s session. Before each block, markers were placed on the bars of *S*'s display designating the positive digits.

TABLE 1

Probabilities of Stimulus Digits in the Positive Sets for the Equal and Unequal Probability Groups

Size of positive set	Group: Equal	Group: Unequal
$s = 1$	4/15	4/15
$s = 2$	2/15, 2/15	3/15, 1/15
($s = 3$)	(1/15, 1/15, 1/15)	(1/15, 1/15, 1/15)
$s = 4$	1/15, 1/15, 1/15, 1/15	2/15, 2/45, 2/45, 2/45

Results

Figure 1 shows the overall Yes–No RT and error functions for both EQ and UN conditions. The slopes for the EQ functions average 40.4 msec/item which is

almost identical to the UN functions which average 38.4 msec/item, and these, in turn, are similar to those reported by Sternberg in his (1967) experiment (mean slope = 35.6 msec/item). Thus, instructing *S* about the probability variations and explicitly manipulating *P*s within the positive set had little effect on the rate and nature of processing. The 30-msec. average difference in intercepts between the EQ and UN groups most likely represents individual differences in the groups: the EQ group was 43 msec. faster at the outset of the experiment, in the $s = 3$ condition, and 26 msec. faster in the $s = 1$ condition. Differences in RTs among the three digit groupings were negligible, $F < 1.00$.

The data for probability adjustments within the positive set are shown in Table 2. An analysis of variance of the positive stimuli of the $s = 2$ and 4 condition of the UN group revealed that the main effects of *s* (2 vs. 4) and *P* (high vs. low) were significant, $F\ (1,\ 23) = 19.40$, $p < .001$, and $F\ (1,\ 23) = 11.85$, $p < .01$, respectively. The $s \times P$ interaction was highly variable from subject to subject and was not significant, $F\ (1,\ 23) = 1.42$. Table 2 shows that the effect of *s* on Yes RTs is too large to be completely attributable to *P*. Three comparisons are relevant here: (*a*) The differences in *P*-values between $s = 2$ and $s = 4$ in the *EQ* condition was only half the difference between high and low *P*s in the $s = 2$ UN condition (.133–.067 = .061 vs. .200–.067 = .133). Yet the former yielded an effect on RT of 77 msec. which was more than twice the effect on RT produced by the latter *P* difference (28 msec.). (*b*) A similar case can be made for the $s = 4$ UN condition. Here the difference in *P* values was .089 (.133–.044) —slightly greater than the difference in *P* values in the EQ condition. Yet, again, the 48-msec. UN difference was less than the effect of *s* in the EQ condition. (*c*) Within the UN conditions, the effect of *s* was clearly greater than the effect of *P*. The RTs to the .133 stimulus in the $s = 4$ condition were slower than the .067 stimulus in the $s = 2$ condition. Thus when *s* and *P* were directly pitted against each other, *s* was the more potent variable.

Is the effect of *s* independent of the effect of *P*? The data were more equivocal on this point but certainly did not provide strong evidence for an interaction between *s* and *P*. Although not statistically reliable, the within-subjects effect of *P* was greater in the $s = 4$ condition than in the $s = 2$ condition, even though the difference (but not the ratio) in *P*-values was greater in the $s = 4$ condition. However, before this suggestion of an interaction between *s* and *P* is given much weight (which would be evidence against a two-stage theory), one must first consider how the interaction could be affected by rescaling the probability variable since the differences in *P*-values for the different set sizes were not taken at the same points on the *P* scale. The variation in *P* for $s = 4$ occupied a lower range of the scale (.044 to .133) than did the variation in *P* for $s = 2$ (.067 to .200). Hence, if one wished to assume a transform between objective and subjec-

TABLE 2

Mean Correct Reaction Times (RTs) for Yes Responses as a Function of Group, Set Size (*s*), and Probability (*P*)

Set size	Group						P-effect (Low *P* minus high *P*)
	Equal		Unequal				
	P	RT (msec.)	High P	RT (msec.)	Low P	RT (msec.)	
2	.133	454	.200	484	.067	512	28
4	.067	531	.133	528	.044	576	48
s effect (4 minus 2)		77		44		64	

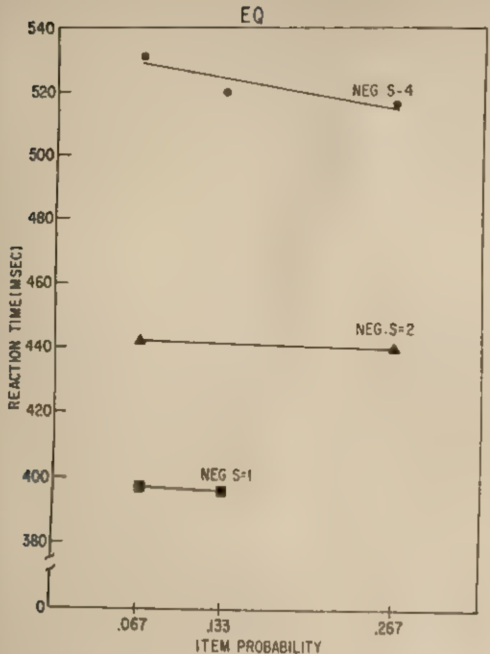


FIGURE 2. Mean negative correct reaction time as a function of probability for the EQ condition.

tive P that expanded the smaller values of the P scale, the interaction between s and P could be eliminated. Such an assumption could be justified on the basis of studies (e.g., Hohle & Gholson, 1968; Miller & Pachella, 1973) that have demonstrated that differences between small P-values produce larger effects on RT than equal differences in larger P-values. It is possible to construct a between-subjects test of the $P \times s$ interaction—which avoids the scaling problem—by comparing the data from the EQ and UN groups. This

TABLE 3

MEAN CORRECT REACTION TIMES (RTs) FOR YES RESPONSES AS A FUNCTION OF SET SIZE (s) AND PROBABILITY (P) WITH CORRECTION FOR GROUP DIFFERENCES

Set size	Probability .133	Probability .067	P-effect (.133 minus .067)
2	480(Equal)	512(Unequal)	32
4	528(Unequal)	557(Equal)	29

Note. A total value of 26 msec., the difference between the equal and unequal groups in the $s = 1$ condition, has been added to the equal RTs.

can be done by equating for the between-group differences on the basis of the common $s = 1$ condition which was 26 msec. faster in the EQ group. Thus if 26 msec. is added to the EQ values as shown in Table 3, then the larger P-effect found in the $s = 4$ group for the within-s analysis is eliminated.

Figures 2 and 3 show the effect on RTs of variations in item probability, set size, and response. Variations of P within the positive set had a considerably greater effect than variations of P within the negative set. The average slope of the functions relating RT to P within the positive set, are 14 times as steep as the corresponding negative functions: -48 msec. per .1 increase in P for the positive slope and -3.4 msec. per .1 increase in P for the negative slope. (By comparison, if the effect of s is described in terms of P, then the resulting slope, -60 msec. per .1 increase in P, is steeper than the function relating RT to P within a given sized positive set.) Probability thus had only a negligible effect on RTs to items within the negative set. In five of the six negative set probability functions, the F ratio for the effect of P was less than 1.00. In the UN group, $s = 2$ condition, there was a significant effect of P: $F(3, 69) = 4.91$, $p < .01$. However, the function for that condition is nonmonotonic and the effect is primarily attributable to the point at $P = .067$ which is *slower* than the point at $P = .044$.

Repetition effects were examined as a function of response type, probability, set size, and group. (Since a number of Ss did not have any entries for repetitions for some of the low probability stimuli in some conditions, the nonrepetition RTs for the corresponding cells for these Ss were not included in the analyses.) Stimulus repetition effects were larger than response repetition effects. Trials in which the stimulus was repeated from the preceding trials resulted in RTs that averaged 25 msec. less than RTs from trials in which the successive stimuli were different but the response was the same. When a different response was required, RTs were 6

msec. longer than when the response was repeated but different stimuli were presented. Table 4 shows that larger stimulus repetition effects were associated with larger set sizes. Response type, *P*, and groups did not yield a consistent effect on the magnitude of the stimulus repetition effect. The magnitude of the response repetition effect did not vary systematically with any of the major experimental variables. Since there were relatively few repetition trials the results shown in Figures 1, 2, and 3 would be only negligibly affected if only nonrepetition trials were used.

Table 5 shows the correct RTs, error RTs, and error rates for the major experimental conditions. Yellot's (1971) correction for fast guessing and the speed-accuracy trade-off applied to these data had only a negligible effect on the comparisons described above.

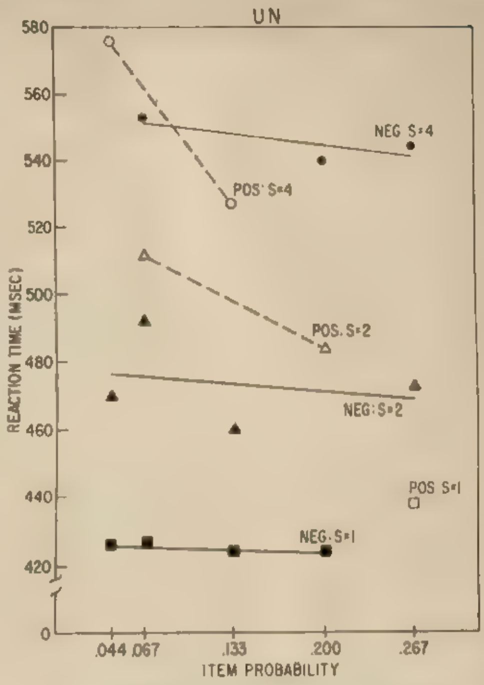


FIGURE 3. Mean correct reaction times as a function of probability and response in the UN condition.

DISCUSSION

Variations in *P* produced by variations in *s* led to considerably greater effects on RT than identical variations in *P* with *s* held constant. Thus it is clear that not all the effect of *s* can be attributed to *P*. The results of this memory scanning experiment are thus consistent with the information conservation studies of Hyman (1963) and Hohle and Gholson (1968) in showing that the effect on RTs of variations in set size of equiprobable stimuli is greater than the effect which is produced by variations of *P* within a given set size. The interaction between *s* and *P* was not significant although there was a trend, within-*s*, for larger *P* effects to be present in the larger positive sets (although this was not the case in the between-*S* comparison).

In the present experiment, if it is assumed that *P* effects reflect the functional set (i.e., the set which is scanned), then the lack of such effects in the negative set would indicate that the negative set was rarely functional (or scanned). What might have occurred in the Kreuger (1970), Theios et al. (1973), and Miller and Pachella (1973) experiments, was that their *S*s, faced with positive and negative sets of equal size, varied the set which they scanned from trial to trial. What determines which set gets scanned? What does the *S* treat as positive? This decision could *not* primarily be based on response probabilities since in the present experiment as well as Sternberg's (1967, 1969) experiments the probability of a positive response was only 4/15. Moreover, variations of response probability independent of stimulus probability have only

TABLE 4

STIMULUS REPETITION EFFECTS (IN MSEC.) AS A FUNCTION OF RESPONSE (POSITIVE OR NEGATIVE), SET SIZE (*s*), AND STIMULUS PROBABILITY

Set	Stimulus probability .044	.067	.133	.200	.267	*M*
Positive						
s = 1					5[a]	5
s = 2		[b]	13	41		27
s = 4	9	93	71			58
Negative						
s = 1	12	0	12	6		−2
s = 2	23	8	21		3	14
s = 4		54	12	50	36	38
M	7	39	26	28	15	

Note. Entries are the difference, in msec., between mean RTs for response nonrepetition and stimulus repetition trials. Positive entries indicate that the repetition trials were of shorter latencies.

[a] Includes both stimulus and response repetition effects.

[b] Insufficient data.

TABLE 5

MEAN CORRECT REACTION TIME (MSEC.), ERROR REACTION TIME (UNDERLINED, IN MSEC.), AND ERROR RATE (IN PARENTHESES) AS A FUNCTION OF GROUP, SET SIZE, AND RESPONSE AND STIMULUS PROBABILITY

Probability	Positive response						Negative response					
	Group equal			Group unequal			Group equal			Group unequal		
	$s = 1$	$s = 2$	$s = 4$	$s = 1$	$s = 2$	$s = 4$	$s = 1$	$s = 2$	$s = 4$	$s = 1$	$s = 2$	$s = 4$
.267	412 324 (.089)			438 364 (.076)				442 456 (.021)	516 586 (.023)		471 525 (.012)	544 508 (.042)
.200					484 408 (.113)					424 548 (.009)		540 637 (.016)
.133		454 410 (.078)				528 584 (.052)	396 490 (.007)		520 521 (.030)	424 400 (.003)	460 471 (.024)	
.067			531 498 (.078)		512 524 (.132)		397 389 (.009)	444 482 (.015)	531 622 (.078)	427 609 (.014)	492 508 (.028)	551 587 (.013)
.044						576 511 (.142)				426 500 (.003)	470 490 (.017)	

slight effects on RT (Biederman & Zachary, 1970). It would appear that the number of items is the most potent determinant of what gets scanned in memory: Left to his own devices, S scans the set with the smaller number of items.[4] It seems reasonable to assume, however, that when positive and negative sets are of equal size, other variables (e.g., probability, repetition, value) would determine which stimuli are scanned on a given trial. Also, if S had insufficient time or memory to determine the complement of a positive set, he might scan the positive set even though it was larger than the negative set.

The interaction between P and Response Type observed in this experiment is contrary to what would have been expected from an encoding theory of stimulus probability as advanced by Klatzky and Smith (1972) and Miller and Pachella (1973). These investigators reported that P did not interact with memory set size but did interact with stimulus degradation. An encoding theory of probability effects might be reconciled with the above findings if it is assumed that prior to the presentation of the probe, S "prepares" only for the positive set items, and the effect of this preparation is on the encoding stage. If the amount of preparation was related to P then larger P effects in the positive than in the negative set might be expected. However, such a theory is inconsistent with the complete absence of any interaction between degradation and response type in the Sternberg (1967) experiment.

If P does not affect encoding, then where does it have its effect? One possibility might be at the response selection stage following memory scanning. If positive set matches are signaled not simply by a match signal but by the digit itself along with a match signal, then response selection could be biased in tune with the probabilities of the positive set stimuli. After all, S knows what the digit was that provided the match. Such a theory would be consistent with the finding that the effects of P and S–R compatibility interact (Sanders, 1970). In fact, with naming tasks (which are of very high compatibility), there is little or no effect of probability (Miller & Pachella, 1973; Theios, 1974). Of the studies discussed above, the only finding that is apparently incompatible with this view is the interaction between P and degradation reported by Miller and Pachella. In that experiment, degradation was produced by the presentation of a light field immediately prior to the presentation of the digits of a CRT display. The field, which reduced the contrast between the digits and background, decayed over time. That is,

[4] Perhaps it is for this reason that it is so difficult to follow the instruction, "Don't think about alligators."

if *S* waited, the digits became clearer. It is possible that such a discriminability manipulation could have induced a response selection strategy which was biased in favor of high-*P* stimuli (assuming that response selection is achieved by a combination of digits and match signals). But even if another method of producing degradation revealed an interaction between *P* and degradation, *P* effects could still be localized at the response selection stage if it was assumed that the *P* bias was to a visual representation of the digit.

REFERENCES

Biederman, I., & Zachary, R. A. Stimulus versus response probability effects in choice reaction time. *Perception & Psychophysics*, 1970, **7**, 189–192.

Briggs, C. E., & Johnsen, A. M. On the nature of central processing in choice reactions. *Memory & Cognition*, 1973, **1**, 91–100.

Briggs, G. E., & Swanson, J. M. Encoding, decoding, and central functions in human information processing. *Journal of Experimental Psychology*, 1970, **86**, 296–308.

Hawkins, H. L., & Hosking, K. Stimulus probability as a determinant of discrete choice reaction time. *Journal of Experimental Psychology*, 1969, **82**, 435–440.

Hick, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 1952, **4**, 11–26.

Hohle, R. H., & Gholson, B. Choice reaction times with equally and unequally probable alternatives. *Journal of Experimental Psychology*, 1968, **78**, 95–98.

Hyman, R. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 1953, **45**, 186–196.

Klatzky, R. L., & Smith, E. E. Stimulus expectancy and retrieval from short-term memory. *Journal of Experimental Psychology*, 1972, **94**, 101–107.

Krueger, L. E. Effect of stimulus probability on two-choice reaction time. *Journal of Experimental Psychology*, 1970, **84**, 377–379.

LaBerge, D., & Tweedy, J. R. Presentation probability and choice time. *Journal of Experimental Psychology*, 1964, **68**, 477–481.

Miller, J. O., & Pachella, R. G. Locus of the stimulus probability effect. *Journal of Experimental Psychology*, 1973, **101**, 227–231.

Sanders, A. F. Some variables affecting the relation between relative stimulus frequency and choice reaction time. *Acta Psychologica*, 1970, **33**, 45–55.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

Sternberg, S. Two operations in character recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, 1969, **30**, 276–315.

Theios, J. Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.), *The psychology of learning and motivation*. Vol. 7. New York: Academic Press, 1974.

Theios, J., Smith, P. G., Haviland, S. E., Traupmann, J., & Moy, M. C. Memory scanning as a serial, self-terminating process. *Journal of Experimental Psychology*, 1973, **97**, 323–336.

Yellott, J. I., Jr. Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, 1971, **8**, 159–199.

(Received November 1, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1108–1115

STIMULUS ENCODING AND DECISION PROCESSES IN RECOGNITION MEMORY [1]

JAMES F. JUOLA,[2] GLEN A. TAYLOR, AND MICHAEL E. YOUNG

University of Kansas

Recognition memory for words and pictures was tested using a fixed, memorized-set procedure (Experiment I) and a continuous procedure (Experiment II). Test items were presented twice each at lags of 4, 12, or 24 intervening items, and any item could be tested using the same or different stimulus form (word or picture) at each lag. A recognition model that assumes successive encoding, decision, and response stages was used as the theoretical framework for interpretation of the results. The analysis indicated that stimulus form and lag affected encoding processes in similar ways for Experiments I and II. Differences were obtained for the decision stage, however, as stimulus form apparently affected decision processes in Experiment II but not in Experiment I.

In binary-classification tests of recognition performance, *S* must categorize a stimulus as being either an *old*, or target stimulus (i.e., one that has been presented before), or a *new* stimulus, or distractor (i.e., one that has not previously been presented). Recognition test procedures are of two principal types: (*a*) the target set consists of a number of items that have been studied prior to the recognition test, or (*b*) the target set consists of all items previously presented during a continuous series of test items.

The former procedure has been used to study both short-term recognition (Sternberg, 1966) and long-term recognition (Juola, Fischler, Wood, & Atkinson, 1971), and Atkinson and Juola (in press) have developed a model to account for the data. The model proposes three processing stages (encoding, decision, and response) that are executed sequentially during recognition. The encoding stage includes analysis of the test stimulus and the construction of an internal representation that is unique to the form and modality of input. This representation is called the perceptual code. Each perceptual code is then mapped onto the particular long-term memory location assigned to various perceptual codes of the same item. This representation in memory is called the conceptual code. The generalized activation level of the information represented by the conceptual code is converted into a familiarity value for the test item. Thus, the encoding stage includes all the processes involved in identifying a test stimulus and retrieving a subjective estimate of its familiarity.

The decision stage of the model operates on the familiarity output by the encoding stage. Familiarities are evaluated with respect to two decision criteria such that values below a low criterion lead to immediate negative responses and values above a high criterion lead to immediate positive responses. If the familiarity falls between the two decision criteria, the response stage is delayed until after a more extensive search of memory has been completed. This search process might be similar to the serial and exhaustive scanning model proposed by Sternberg (1966, 1969). The major difference between the two models is that in Sternberg's representation a decision to respond is made only after an exhaustive search of the target set has been completed, whereas in the model described above, a decision can be made either before or after this search. A decision made on the basis of familiarity alone is rapid, but error prone, and a

[1] This research was supported in part by funds from Biomedical Sciences Support Grant RR-07037 from the National Institutes of Health, University of Kansas General Research Grant 3294-5038, and Grant MH-24637-01 from the National Institute of Mental Health. The authors thank Greg Cermak, George Kellas, Roy Lachman, and Carl McFarland for comments on an earlier draft of this paper.

[2] Requests for reprints should be sent to James F. Juola, Department of Psychology, University of Kansas, Lawrence, Kansas 66044.

decision made after the extended search is slower, but accurate.

One of the main purposes of the present study was to extend the recognition model to account for data from continuous recognition experiments. In the continuous task, a series of items is presented, and S is instructed to make negative responses to all items on their initial presentations and to make positive responses to all repeated items (e.g., Shepard & Teghtsoonian, 1961). Studies by Hintzman (1969) and Okada (1971) have shown that both the number of presentations and the lag between successive presentations affect performance. In general, repeated tests result in faster positive responses and fewer errors, with mean positive response latency increasing as a nonlinear function of lag between tests. These results cannot be accommodated by the Sternberg model, probably because the target items generally are not retrievable as a set which could be scanned for the presence or absence of the test item (Okada, 1971).

Similar effects of repetitions and lags have been observed in studies using the memorized target set procedure (Fischler & Juola, 1971; Juola et al., 1971). Repetitions have been shown to reduce both the mean error rate and mean response latency for target items, whereas the error rate increases and latencies show either a slight increase or decrease for repeated distractors. These results have been interpreted in terms of the model by assuming that repetitions facilitate the encoding stage and increase mean familiarity values for both targets and distractors (Juola, 1973). For target items, the proportion of the familiarity distribution below the low criterion is assumed to be equal to the miss rate. The proportion between the two criteria and that above the high criterion are assumed to be equal to the proportions of slow and fast positive decisions, respectively. Thus, target repetitions should result in fewer errors, fewer slow decisions, and more fast decisions (and therefore a faster mean execution time for correct positive decisions). For distractors, the proportion of the familiarity distribution above the high criterion is assumed to be equal to the false alarm rate. The proportions between the two criteria and below the low criterion are assumed to be equal to the proportions of slow and fast negative decisions, respectively. Thus, repetitions of distractor items should result in more errors and a slower mean execution time for correct negative decisions. Repetition effects on the encoding and decision stages should be additive to produce faster mean response times for targets. For repeated distractors, latencies could be either greater than or less than the times for responses to distractors on their first tests, depending on the relative changes in mean execution times for the encoding and decision stages.

The analysis of repetition effects on negative and positive responses has proved useful in separating stimulus encoding and decision effects on recognition performance. Another technique that has been used for this purpose involves the use of different stimuli to represent a given test item. For example, Juola (1973) studied repetition effects for items tested twice using the same or different visual stimuli (words or pictures) on both tests of targets and distractors. The results indicated that the encoding stage was executed more rapidly for all repeated items; however, this effect was much larger for those items tested with the same stimulus form on both presentations. The decision stage was apparently independent of the stimulus forms used.

In the present study, a similar manipulation of stimulus forms was used to help in isolating encoding and decision processes and their relative contributions to recognition performance. In Experiment I Ss were tested using the fixed target set procedure. Targets and distractors were presented twice each at three different lag conditions, and either the same or different stimulus forms were used at all lags. Experiment II included the same manipulations as Experiment I in a continuous recognition task.

Method

Experiment I

Subjects. The *S*s were 50 right-handed, female undergraduates who participated for credit in an introductory psychology class. During the experiment, 3 *S*s were dismissed for reported difficulty in identifying the visual stimuli, and the data for 7 additional *S*s were discarded for failure to meet an arbitrary error criterion of 10%. The data are thus based of an *N* of 40.

Materials. The stimuli were selected from a pool of 96 different items, each of which was represented by a word and a picture. The words were common one- or two-syllable nouns, three to eight letters in length, and the pictures were easily identifiable, outline drawings of the objects named by the nouns. The stimuli were placed on 6 × 9 in. white cards for presentation in an Iconix tachistoscope. The words were typed in Orator typeface with an IBM Selectric typewriter, and they varied in length from .8 cm. to 2.0 cm. Xerox copies of the original drawings were affixed to cards, and they varied between .9 cm. and 3.8 cm. in diameter. When placed into the tachistoscope, the stimuli appeared about 1.0 cm. (i.e., about 36 min. of visual angle from *S*'s viewpoint) below a central, dark fixation point.

Design. The 96 items were randomly divided into two sets of 48 items each. The first set was used to generate target lists for 20 *S*s, and the remaining 48 items were used as distractors. For a second group of 20 *S*s, the target and distractor sets were reversed so that for the experiment as a whole, all items were used equally often as targets and distractors. Both target sets were randomly ordered into two different lists to provide the target lists used for four groups of 10 *S*s each.

The test session consisted of 192 trials during which all items were tested twice. The *S*s were randomly assigned to three different test sequences. In these sequences, item presentation order was randomly determined with the constraint that either 4, 12, or 24 intervening trials occurred between successive presentations of the same target or distractor items. Sixteen targets and 16 distractors were repeated at each of these three lag intervals, which, after the first 24 trials, were distributed uniformly throughout the test sequences. Half of the target and distractor items were tested as words (W trials) and half were tested as pictures (P trials) on their first presentations. At each of the three lag conditions for repeated targets and distractors, four items previously tested as words were tested again as words on their second presentations (WW trials), four items tested as words were retested as pictures (WP trials), four items tested as pictures were retested as words (PW trials), and four items were tested as pictures on both presentations (PP trials).

Procedure. Each *S* was contacted about 24 hr. before the experimental session and was given a target list of 48 nouns. The *S*s were told to memorize the list in serial order before being tested on the following day. The experimental session began with a written, serial recall test for the target list. M[...] *S*s correctly recalled all the target words; howe[...] those who made errors were retested until the wo[...] were recalled in correct serial order.

The *S* was then seated in front of the tachist[...] scope, and the exact nature of the test procedur[...] was explained. A series of 20 practice trials on [...] modified Sternberg task was run to familiarize [...] with the procedure. For these trials, *S* was tol[...] to make a positive response if a digit from 0 to 4 [...] presented and to make a negative response i[...] test digit was 5–9. All 10 digits were presented [...] in a random order.

The procedure for the practice trials was identi[...] to that for the experimental test sequence. [...] trial started when *E* placed a test stimulus [...] into the tachistoscope and gave *S* a ready sign[...] The *S* then depressed a foot switch that initiate[...] a millisecond timer. The tachistoscope was pr[...] grammed to illuminate a preexposure field [...] dark fixation point 10 msec. after *S* pressed the foot switch. This field was maintained for 490 [...] and was followed immediately by a 350-ms[...] posure of the test stimulus. The field rem[...] dark until the next trial. The *S* was instructed [...] rest her hands on a box upon which two 3.2-cm circular buttons were mounted about 9.0 cm. apart. When either button was pressed, it activated a microswitch that stopped the timer. All *S*s were instructed to indicate a positive response to target items by pressing the right-hand button and to indicate a negative response to distractors by pressing the left-hand button.

The practice trials were followed by a brief res[...] period during which *S* was told that targets and distractors could be tested either as words or as pictures, and she was instructed to respond as rapidly as possible on each trial while being careful to avoid making errors. During the test session no feedback was provided for correct responses, but *S*s were informed of all errors. The session averaged about 45 min. in length during which *S*s we[...] allowed to take a maximum of three rest periods at irregular intervals.

Experiment II

Subjects. The *S*s were 22 right-handed female undergraduates who participated for credit in an introductory psychology class. During the experiment 1 *S* was dismissed for reported difficulty in identifying the visual stimuli, and the data for an additional *S* were discarded for failure to meet an arbitrary error criterion of 10%. The data are thus based on an *N* of 20.

Materials. The stimuli and apparatus were the same as those used in Experiment I.

Design. The test sequences used in Experiment II were identical to those used in the previous study. The sequences were randomly assigned to *S*s.

Procedure. The test session began with a description of the testing procedure. A series of 20 trials using the digits 0–9 as stimuli was run to

TABLE 1

MEAN LATENCIES AND STANDARD ERRORS OF THE MEANS (IN MSEC.) FOR CORRECT RESPONSES AND MEAN ERROR PERCENTAGES FOR ALL CONDITIONS OF EXPERIMENTS I AND II

Measure	Experiment I								Experiment II			
	Targets				Distractors							
	First presentations											
	W		P		W		P		W		P	
M	868		1,048		901		1,228		940		1,003	
SE	19.2		32.2		24.7		44.5		33.8		42.3	
% errors	13.1		11.2		1.9		5.1		4.8		7.1	
	Second presentations											
	WW	WP	PW	PP	WW	WP	PW	PP	WW	WP	PW	PP
Lag 4												
M	680	803	702	654	760	902	846	730	751	820	858	718
SE	16.6	25.3	15.5	17.8	19.0	26.9	29.5	17.2	18.5	29.3	25.3	23.0
% errors	3.1	0.0	2.5	1.5	4.8	7.1	3.5	3.3	2.9	3.9	.7	0.0
Lag 12												
M	708	835	713	707	836	1,050	884	813	831	879	871	750
SE	18.5	23.7	20.6	22.2	24.6	44.9	34.4	25.9	31.6	29.2	36.0	24.4
% errors	1.5	1.7	1.7	3.3	7.4	7.7	2.5	7.5	3.7	7.3	3.0	1.2
Lag 24												
M	694	894	769	761	812	1,097	863	858	889	921	894	778
SE	16.9	29.0	25.1	25.9	20.5	47.2	26.4	24.6	34.1	35.8	26.9	23.5
% errors	.6	4.8	.6	3.1	2.5	3.8	2.9	2.3	2.7	10.4	5.1	0.0

Note. Abbreviations: W = items tested as words, P = items tested as pictures, WW = items tested as words on both presentations, WP = items tested first as words and retested as pictures, PW = items tested first as pictures and retested as words, and PP = items tested as pictures on both presentations.

familiarize *S* with the task. For these trials, *S* was told to make a negative response to any digit on its first presentation and a positive response to any repeated digit. All 10 digits were presented twice in a random order. All other details of the practice trials and the actual test session were identical to those of Experiment I. The only difference for the present study was that, rather than making positive responses to items belonging to a previously memorized set, *S*s were instructed to make positive responses to any items previously presented during the test session, regardless of whether or not the actual stimulus form (W or P) was the same on both tests of an item.

RESULTS

The mean response latencies,[3] standard errors of the mean latencies, and mean proportions of errors for all conditions of Experiments I and II are presented in Table 1. The data presented here are from Trials 25 to 185 or 186 only (depending on the specific test sequence), over which the distributions of all types of trials were approximately uniform.

In both experiments mean response times were slower to pictures than to words for items on their initial presentations. (This trend was obtained for 37 of 40 *S*s for targets and 38 of 40 *S*s for distractors in Experiment I and for 18 of 20 *S*s in Experiment II.) The latency data for repeated items were analyzed separately for targets and for distractors in Experiment I and for Experiment II. Three $2 \times 2 \times 3$ repeated measures analyses of variance were run, with factors corresponding to the stimulus form (W or P) used on the first presentation of the test item, the stimulus form used on the second test, and the lag between presentations, respectively.

[3] The latency data were based on 89.7% of the total trials run in Experiment I and 92.2% of the trials in Experiment II. The data excluded were for Experiment I: 5.8% *S* errors, 3.6% *E* errors, and .9% "long" responses (eliminated using the correction technique described by Wescourt & Atkinson, 1973); for Experiment II: 4.9% *S* errors, 2.8% *E* errors, and .1% long responses.

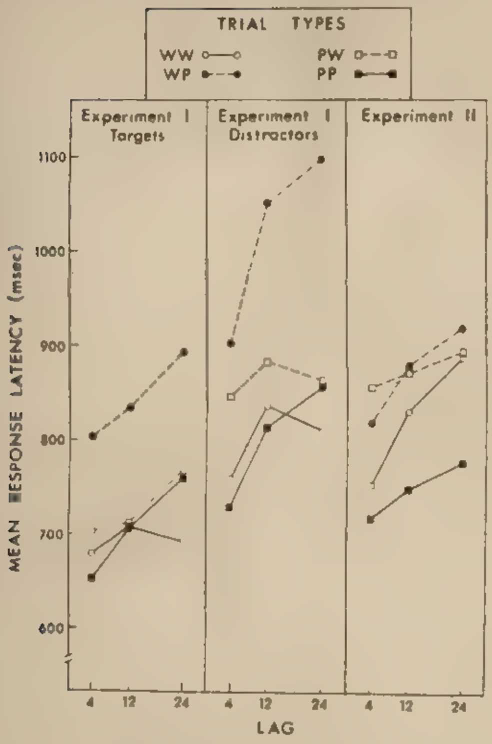


FIGURE 1. Mean response latencies for four types of items on second test presentations as functions of lag between first and second presentations.

All main effects of stimulus form were significant. In Experiment I, responses to repeated targets were faster if they had been tested as pictures on their first tests, $F\ (1, 39) = 34.21$, $MS_e = 9{,}375$, $p < .01$, and as words on their second tests, $F\ (1, 39) = 60.69$, $MS_e = 8{,}223$, $p < .01$. Responses to repeated distractors were also faster if they had been tested first as pictures, $F\ (1, 39) = 31.48$, $MS_e = 22{,}659$, $p < .01$, and then as words, $F\ (1, 39) = 23.48$, $MS_e = 28{,}796$, $p < .01$. In Experiment II, response latencies were shorter for repeated items tested first as pictures, $F\ (1, 19) = 9.51$, $MS_e = 8{,}940$, $p < .01$, and responses were faster for items tested as pictures on their second presentations, $F\ (1, 19) = 17.60$, $MS_e = 5{,}042$, $p < .01$. In general, the response time data indicated that latencies were shorter for repeated items if the same stimulus forms were used on both tests than if they differed. This conclusion was supported by the significant interactions between the forms use[illegible] the first and second tests, $F\ (1, 39) =$ [illegible] 7.89, $MS_e = 9{,}919$, $p < .01$, for ta[illegible] in Experiment I; $F\ (1, 39) =$ [illegible] $= 33{,}317$, $p < .01$, for distra[illegible] periment I; and $F\ (1, 19) =$ [illegible] $= 8{,}872$, $p < .01$, in Experime[illegible]

The main effects of lag were [illegible] cant, indicating that response [illegible] fo[illegible] repeated items increased as a f[illegible] the number of intervening test[illegible] the first and second tests of [illegible] $F\ (2, 78) = 19.09$, $MS_e = 10{,}190$, [illegible] for targets in Experiment I; F (2[illegible] 30.30, $MS_e = 15{,}015$, $p < .01$, fo[illegible] tractors in Experiment I; and [illegible] $= 39.42$, $MS_e = 3{,}508$, $p <$ .0[illegible] tractors in Experiment II. In [illegible] I, lag interacted with the stimu[illegible] used on the second tests of targets, F (2, [illegible] $= 3.82$, $MS_e = 8{,}977$, $p < .05$, and di[illegible] tractors, $F\ (2, 78) = 6.39$, $MS_e =$ [illegible]20, $p < .01$. In Experiment II, how[illegible] interacted with the stimulus form[illegible] on the first tests of repeated items, F (2, [illegible] $= 5.30$, $MS_e = 4{,}848$, $p < .01$. None [illegible] the other interactions approa[illegible] cance. The interaction effects of [illegible] and stimulus forms are apparent in [illegible] where, for Experiment I, the W[illegible] functions showed a large and nea[illegible] increase across lags whereas the [illegible] and PW functions increased only slig[illegible] In Experiment II, the WW and W[illegible] exhibited the greatest lag effe[illegible] the PW and PP functions showed [illegible] nearly parallel increase.

A comparison of the latency data [illegible] initial vs. repeated presentations (av[illegible] across lags) in Experiment I showed that responses were faster on second tests than first tests for at least 35 of 40 *Ss* in all conditions but one. The exception was for distractor items tested first as words then as pictures, for which 32 of 40 *Ss* were slower on the second tests than on initial presentations. Although a similar analysis of the data from Experiment II involves a confounding of response type (positive or negative) with presentation number, it was noted that at least 36 of 40 *Ss* in all conditions were faster on the second tests.

The error data for initial presentations of test items were analyzed using Wilcoxon matched-pairs signed-ranks tests. The difference between the mean proportions of errors to words and pictures was not significant for the target data from Experiment I; however, significantly more errors were made for distractor pictures than words ($p < .01$, two-tailed test). The error rate was also higher for picture stimuli than for words on initial tests in Experiment II ($p < .05$, two-tailed test).

The error data for second test presentations were analyzed separately for the four trial types collapsed across lags and for the lag data collapsed across trial types. A Friedman two-way analysis of variance by ranks indicated that there were no significant differences between the four trial types for the target and distractor data of Experiment I. For Experiment II, however, the Friedman analysis was significant, χ^2 (3) = 9.6, $p < .05$. Multiple sign tests indicated that the only significant difference was between WP trials (mean error rate = 7.1%) and PP trials (.5%).

Although the error rate increased as a function of lag for targets in both Experiments I and II, neither increase was significant. A Friedman analysis showed that lag had a significant effect on the error rate for distractors in Experiment I, χ^2 (2) = 9.3, $p < .01$, and multiple sign tests indicated that fewer errors were made at Lag 24 (2.9%) than at Lag 4 (4.4%) or at Lag 12 (6.6%).

A comparison of the error data for initial vs. second presentations (averaged across lags) in Experiment I showed that the error rate decreased with repetitions for all target conditions (Wilcoxon test, $p < .01$). For distractors, the error rate increased from Presentation 1 to Presentation 2 for those items tested first as words ($p < .01$), but there was no significant difference in the error rate for P trials vs. PW and PP trials. For Experiment II there was no significant difference in the error rate for W trials vs. WW and WP trials, but the error rate was less for PW and PP trials than for P trials ($p < .01$).

Discussion

The present studies were designed to investigate stimulus form and lag effects on recognition performance and to test the generality of a recognition model. The results will be interpreted in terms of the model in that differences in error rates for different target or distractor conditions will be assumed to represent differences in the relative familiarities of those items. The assumption that errors are generated only from the decision stage will necessarily have to be modified if evidence exists that errors can result from failures in other processing stages, such as in stimulus encoding or response selection. On the other hand, differences in response latencies that are not accompanied by differences in error rates will be assumed to represent differential facilitation of encoding processes.

In both experiments, responses were faster to word stimuli than to pictures on initial tests. This is to be expected since the pictures were original drawings that had never before been seen by *Ss*. According to the model, these differences should reflect greater expected durations for the encoding stage on P trials than on W trials. The decision stage should not be affected by stimulus form since the familiarity value of any test item is determined by information derived from the conceptual code onto which the perceptual codes of the word and picture are mapped. This interpretation is supported by the error data from initial target presentations in Experiment I which showed no significant differences on W and P trials. For initial tests of distractors in both Experiments I and II, however, significantly more errors were made on P trials. These results could be taken to indicate that distractor pictures result in different familiarity values than distractor words. However, it is also possible that, as for target items, they are about equally familiar and the error rate difference is due to a greater probability of misidentifying the pictures than the words. This effect could have been mitigated for targets since *Ss* had previously studied the names of the pictures.

Responses were faster in both experiments for repeated items that were tested twice using the same stimulus form (i.e., for WW and PP trials) than if the forms were different (i.e., for WP and PW trials). In Experiment I, there were no error rate differences between these four types of trials, but stimulus forms did affect the error rates for repeated items in

Experiment II. The Experiment I results are consistent with other data (Juola, 1973) and with the interpretation that the decision to respond before or after an extensive memory search is based only on the familiarity value derived from the conceptual code of the test item. Although repetitions might facilitate execution of the encoding stage, especially if the stimulus forms are also repeated, this facilitation apparently ends once the appropriate conceptual code is uniquely identified. The decision process in recognition tasks employing the fixed target set procedure appears to be independent of the stimulus form used at the time of test or during earlier presentations of the test item.

The error data for repeated items in Experiment II indicate that memory for stimulus form might be used in the decision process. This decision could occur in at least two levels of processing: (*a*) during the construction of the perceptual code of the test item or (*b*) during extended memory search if the familiarity derived from the conceptual code is of an intermediate value. In either case a decision could be based on obtaining a match between the perceptual codes constructed on two presentations of an item. Thus for the continuous paradigm, it is possible that the decision stage is not separable from the encoding stage of recognition.

An analysis of the lag data also supports the contention that encoding and decision processes are separable in Experiment I and confounded in Experiment II. In Experiment I, response latencies for all types of trials increased with lag, but this increase was more pronounced for items presented as pictures on their second tests. Lag effects for repeated items were apparently independent of the forms used on the initial tests of target and distractor items. The lack of any effects of stimulus form of initial items on the lag data suggests that the stimulus form effects for repeated items were due solely to differences in the encoding stage. The larger lag effects for WP and PP trials than for WW and PW trials indicate that not only is the encoding stage for picture stimuli affected more by repetitions than that for words (e.g., the repetition effect was much larger for PP trials than for WW trials), but also the facilitative effect of repetitions decays more rapidly for items tested as pictures.

In Experiment II, increasing lags between presentations also caused increases in response latencies for repeated items. However, the pattern of interactions between stimulus form and lag effects differed from Experiment I in that the stimulus forms used on the initial tests, but not those used on second tests, interacted with lag. The lag effects on response latencies for repeated items were much greater for those tested as words on their initial presentations than for those tested as pictures. This result is consistent with the hypothesis that *S*s compare the perceptual code of the test item with retained visual information from earlier tests of the item. For WW and PP trials, a decision can be made during or immediately after the construction of the test item's perceptual code. For WP and PW trials, this decision cannot be made until the alternative perceptual code is activated (or reconstructed) which might first necessitate access of the common conceptual code. This representation of the decision process is predicated on two assumptions: (*a*) visual information for both words and pictures is retained in long-term memory (Hintzman & Summers, 1973) and (*b*) the visual information for words decays faster (or becomes nondiscriminative sooner) than that for pictures (Snodgrass, Wasser, Finkelstein, & Goldberg, 1973).

The model derived from earlier recognition studies using the fixed memory set paradigm is consistent with the data from Experiment I, but it needs to be modified to handle the Experiment II results. For the case in which the target items are members of a well-defined set, the decision to make a positive or negative response cannot be based upon matching a current perceptual code with the trace of one presented earlier, since both targets and distractors can be repeatedly tested. For the continuous procedure, however, decisions can be based upon matches obtained during the encoding stage of recognition. In either experimental paradigm, a decision can be determined by the familiarity (or activation level) of the conceptual code that represents the test item. If, however, the familiarity is of some intermediate value that does not allow for relative certainty of discrimination between targets and distractors, a more extensive search of memory might be executed before a decision is made. This search could involve the retrieval of the target set and the scanning of its contents for the presence or absence of the test item if the set is well-learned, whereas, for the continuous paradigm, this search might involve the retrieval of alternative per-

ceptual representations of the test item in a search for one that had been constructed earlier.

REFERENCES

ATKINSON, R. C., & JUOLA, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

ATKINSON, R. C., & JUOLA, J. F. Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*. San Francisco: Freeman, in press.

FISCHLER, I., & JUOLA, J. F. Effects of repeated tests on recognition time for information in long-term memory. *Journal of Experimental Psychology*, 1971, **91**, 54–58.

HINTZMAN, D. L. Recognition time: Effects of recency, frequency, and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, **79**, 192–194.

HINTZMAN, D. L., & SUMMERS, J. J. Long-term visual traces of visually presented words. *Bulletin of the Psychonomic Society*, 1973, **1**, 325–327.

JUOLA, J. F. Repetition and laterality effects on recognition memory for words and pictures. *Memory & Cognition*, 1973, **1**, 183–192.

JUOLA, J. F., FISCHLER, I., WOOD, C. T., & ATKINSON, R. C. Recognition time for information stored in long-term memory. *Perception & Psychophysics*, 1971, **10**, 8–14.

OKADA, R. Decision latencies in short-term recognition memory. *Journal of Experimental Psychology*, 1971, **90**, 27–32.

SHEPARD, R. N., & TEGHTSOONIAN, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, **62**, 302–309.

SNODGRASS, J. G., WASSER, B., FINKELSTEIN, M., & GOLDBERG, L. B. On the fate of visual and verbal memory codes for pictures and words: Evidence for a dual coding system in recognition memory. Paper presented at the Indiana Conference on Theoretical and Mathematical Psychology, Indiana University, April 1973.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

STERNBERG, S. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

WESCOURT, K. T., & ATKINSON, R. C. Scanning for information in long- and short-term memory. *Journal of Experimental Psychology*, 1973, **98**, 95–101.

(Received November 16, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1116–1122

RETRIEVING ATTRIBUTE AND NAME INFORMATION FROM SEMANTIC MEMORY[1]

ELIZABETH F. LOFTUS[2] AND WILLIAM COLE

University of Washington

In two experiments, Ss retrieved instances of well-learned semantic categories (e.g. ANIMALS). In Experiment I, two types of restrictors were used: Sometimes the instance had to possess a certain attribute (e.g., it had to be small), and sometimes its name was restricted (e.g., it had to begin with the letter *M*). Prior knowledge about the type of restrictor differentially influenced the retrieval of these two types of items. In Experiment II, Ss produced category instances that simultaneously satisfied both an attribute and a name restriction (e.g., ANIMAL–SMALL–M). Ss invariably took longer to respond when the name restrictor was presented before the attribute restrictor rather than afterward. A dictionary–network model, which embodies the notion that intrinsically different processing strategies operate in the retrieval of attribute vs. name information, accounts nicely for these results.

One technique for studying semantic memory is to ask *S*s simple questions about material they already know and to measure how long it takes them to answer these questions. The reaction times (RTs) obtained are then used to make inferences about the processes used for retrieval.

In several experiments that have used this technique, *S* was required to produce a member of a category that satisfied some restriction. For example, he might be asked for the name of a *fruit* that is *yellow*, or an *animal* that begins with the letter *Z* (e.g., Freedman & Loftus, 1971; Grober & Loftus, in press; Loftus & Suppes, 1972). In the Freedman and Loftus study, category–adjective and category–letter pairings (hereafter referred to as adjective and letter stimuli, respectively) were randomly "mixed" together in a long sequence of presentations. In the Grober and Loftus study, this mixed condition was included, but, in addition, *S*s received two nonmixed conditions: one in which they saw only adjective stimuli, and the other in which they saw only letter stimuli. It was thus possible to compare the speed of producing a *fruit–yellow* when it occurred in a mixed vs. a nonmixed condition. A major finding was that RT to an adjective stimulus w[as] independent of whether *S* knew that h[e] would be seeing an adjective; that is to say, RT to an adjective was equal in mixed and nonmixed conditions. However, *S*'s responses to letter stimuli were dramatically affected by his expectations; he responded significantly faster in the mixed condition. Because mixing had a differential effect on adjective and letter stimuli, two retrieval processes were thought to occur; that is, the process of retrieving a category member when the restriction is on a *quality* or *attribute* of that member (e.g., it must be yellow) was thought to be different from the process of retrieving a member when the restriction is on the name of that member (e.g., the name must begin with the letter *Z*).

An alternative interpretation was proposed to us by a grant review panel at the National Institute of Mental Health. The panel argued that the adjective–letter difference might not be the result of intrinsically different processing strategies, but rather of the fact that the letter and adjective cues impose different degrees of constraint or uncertainty. When *S* is in a nonmixed–letter condition, he first sees a category for some interval of time, and then a letter appears. He must respond as

[1] This research was supported by a grant from the National Institute of Mental Health. The authors are grateful to Earl B. Hunt, Colin MacLeod, and Steve Poltrock for comments and to Elaine Botti for running the Ss in Experiment II.

[2] Requests for reprints should be sent to Elizabeth F. Loftus, Department of Psychology, University of Washington, Seattle, Washington 98195.

quickly as possible with a member that begins with that letter. During the interval between the category and letter, there is a fixed number of stimulus alternatives that *S* can expect as the forthcoming restricting letter. When *S* is in the nonmixed–adjective condition, however, the set of alternatives he can expect is ill-defined and innumerable. In a mixed condition, there is a great deal of uncertainty as to the identity of the forthcoming restriction; any letter or adjective is possible. The reason that "unmixing" reduces RT for the letter stimuli may be that uncertainty is considerably reduced; the reason it has no effect for adjective stimuli may be that uncertainty is hardly reduced at all.

Experiment I was designed to test the feasibility of the "uncertainty" interpretation. As in earlier studies, *S*s saw nouns paired with adjectives or letters that were presented in mixed and nonmixed conditions. However, sometimes *S* was told at the beginning of the trial that the forthcoming restrictor would be one of two or one of four possibilities. Compared to a condition in which no information is given at the start of the trial, both the two- and four-cue conditions considerably reduce the uncertainty as to the identity of the forthcoming restrictor.

Experiment I

Method

Subjects. The *S*s were 24 undergraduates at the University of Washington. Each participated in one experimental session lasting approximately 45 min.

Materials. Fifty-four noun categories were selected, and each category was paired once with the first letter of a common category member (e.g., FRUIT–P) and once with an adjective descriptive of some common category member (e.g., FRUIT–YELLOW). Thus, 108 unique pairs were created, and each pair was hand printed in block letters on a 5 × 8 in. index card. Each card, then, contained a noun category on the left side and a restricting letter or adjective on the right.

The 108 items were presented in nine blocks of trials. In three blocks, adjective (M–A) and letter (M–L) stimuli were randomly mixed together. Of the remaining six blocks, half contained only adjective stimuli (O–A), and the other half contained only letter stimuli (O–L). Within each type of block (mixed, O–A, or O–L), *S*s received before each trial either zero, two, or four cues, among which was the restrictor being used on that trial. For example, in the two-cue O–L condition, just before the presentation of FRUIT–P, *S* might be told that on this trial the restrictor would be either A or P. The *S* was then shown the category, FRUIT, and after a 1-sec. interval, one of the two cues mentioned was shown. The *S*'s task was to produce from memory an item that belonged to the noun category and began with the restricting letter. On the four-cue trials, before each stimulus was displayed, *S* was given four items, among which was the forthcoming restrictor. On the zero-cue trials, the stimulus cards were displayed without preliminaries.

The three cue conditions (zero, two, or four) and three types of blocks (mixed, O–A, and O–L) resulted in nine distinct blocks of trials. A Latin square balancing procedure determined the ordering of the nine blocks that each *S* received.

Each O–A and each O–L block contained nine adjective and nine letter stimuli, respectively, while the mixed blocks contained a random selection of nine letter and nine adjective stimuli. A new randomization was performed for each *S* with the restriction that throughout the experiment as a whole, a stimulus, such as FRUIT–P, appeared an equal number of times in each of the three cue conditions and as often in a mixed condition as a nonmixed one.

Procedure. Each *S* was told that we were investigating typical memory processes, that he would see category names followed by either an adjective or a letter, and that he was to produce an item that belonged to the category and fit the restriction imposed. The cue conditions were explained, and each *S* was informed before each block of trials which condition was in effect for the block. Sixteen warm-up trials preceded the session, illustrating the varieties of conditions in the experiment.

The *S* sat in front of a screen in which was a window covered by half-silvered glass. The index card containing the stimulus was placed in the dark enclosure behind the mirror and was revealed by illuminating the enclosure. The *S* spoke his response into a microphone placed in front of him.

A trial consisted of the following: (*a*) in two- and four-cue conditions, the cues were spoken by *E*. (*b*) a warning tone was sounded and, following a brief interval, the noun category was illuminated; (*c*) after a 1-sec. interval, the restrictor was automatically illuminated, and simultaneously an electric timer with a dc clutch was started; and (*d*) *S*'s verbal response activated a voice key that stopped the clock and terminated the illumination of the stimulus. If *S* did not respond in 15 sec., the trial was terminated by *E*.

Results

Only correct responses (97%) are included in the following statistical analyses. Responses were grouped into 12 classifica-

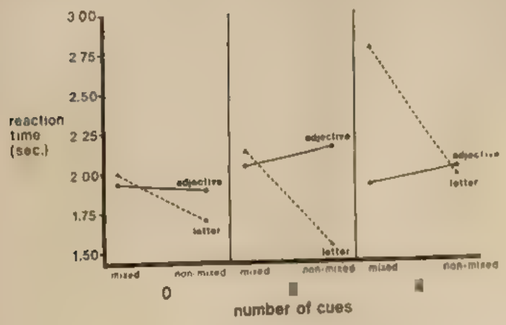


FIGURE 1. Mean reaction times for adjective and letter stimuli in mixed and nonmixed conditions. Data are shown separately for the zero-, two-, and four-cue conditions.

tions depending on type of restrictor (adjective or letter), number of cues (zero, two, or four) and presentation sequence (mixed or nonmixed). Median latencies were obtained for each S's responses in each of the 12 classifications. For each classification, group mean latencies were obtained by averaging medians for individual Ss, and these are plotted in Figure 1. A three-way analysis of variance with repeated measures was done on the latency data. The significant findings were as follows: a nonmixed presentation led to faster responding than a mixed presentation (1.86 vs. 2.12 sec.), F (1, 23) = 6.57, $p < .05$, and this variable interacted with type of restrictor, F (1, 23) = 13.47, $p < .01$. From Figure 1 we can see that the interaction results from the fact that mixing had no effect on adjective trials, but had a substantial and consistent effect on letter trials. Apparently, Ss disregard the cues on adjective trials, but not on letter trials.

Although Ss were somewhat faster when given no cues than when given two or four cues (with zero cues, $\bar{RT}$ = 1.88 sec.; with two, 1.95 sec.; with four, 2.15 sec.) this effect was not statistically significant, F (2, 46) = 1.85, $p > .10$. Furthermore, number of cues did not interact with any of the other factors ($p > .10$).

Discussion

In Experiment I, Ss were asked to retrieve category members that were either (*a*) restricted in that they had to possess a certain attribute, or (*b*) restricted in that their *name* had to possess a certain feature. The RT for producing the first type of member was independent of whether S knew in advance that the restriction would be on the attribute. However, the RT for producing the second type of member was reduced when S knew in advance that the restriction would be on the name. Reducing the uncertainty as to the identity of the forthcoming restrictor with advance cues did not facilitate RT, as an uncertainty interpretation would predict, but in fact slightly increased RT. These data argue against the proposal that the differential effect on adjective and letter stimuli of mixing trials is due to differential uncertainty, and support the notion that intrinsically different processing strategies may be operating.

Although the number of cues had no statistically significant effect in Experiment I, one cannot help but notice that the RT's in the four-cue letter conditions were unusually long. One tentative hypothesis (suggested to us by E. B. Hunt) is that S pays attention to the cues, yet interference is produced by an "overloading" of short-term memory. When adjective stimuli are presented, S acts as if he is ignoring the cues. Confirmation of this possibility must await the outcome of further research. Whatever the explanation, the fact that providing four cues prior to a stimulus has differential effects on the responses to adjective and letter stimuli strengthens o view that intrinsically different processin strategies are operating.

Before suggesting what strategies may be operating in the retrieval of attribute vs name information, we describe a second e periment that bears on the uncertainty iss In this experiment, Ss were presented with stimuli consisting of a noun category plus both an adjective and a letter (e.g., ANIMAL–SMALL–M) and had to produce a member of the category that satisfied the two restrictions. That is to say, the response had to be a member of the category that began with the given letter and to which the adjective was applicable (e.g., MOUSE). The Ss saw the category first, but the order in which the adjective and letter restrictors were presented was varied: on half of the trials the adjective came .5 sec. before the letter (e.g., ANIMAL–SMALL–M), while on the remaining trials the letter came first (e.g., ANIMAL–M–SMALL). Reaction time was taken from the onset of the last restrictor.

When the to-be-presented stimulus is ANIMAL–SMALL–M, during the interval between

SMALL and M there is a fixed number of stimulus alternatives that *S* can expect as the forthcoming restrictor. However, when the letter comes first, and *S* is anticipating the adjective, the set of alternatives he can expect is ill-defined and innumerable. In other words, there is a great deal of uncertainty as to the identity of the forthcoming restrictor—any adjective is possible. An uncertainty hypothesis predicts *S* would respond more quickly in the former case than in the latter.

Experiment II

Method

Subjects. The *S*s were 16 students at the New School for Social Research. Each participated in one experimental session lasting approximately 50 min.

Materials. Forty-six noun categories were selected, and each category was paired with a letter and an adjective (e.g., ANIMAL–SMALL–M). Two index cards were prepared for each category: One contained the adjective restrictor on the left side and the letter restrictor on the right, while the other card contained the reverse arrangement. Throughout the experiment, then, each category was presented as often with the adjective before the letter as with the adjective after the letter.

Procedure. Each *S* was told that we were investigating human memory. He was told he would be presented with a category name followed by a letter and an adjective, and that he was to provide a word that belonged in the overlap defined by the triplet. For example, if shown the stimulus ANIMAL–SMALL–M, he was to produce the name of a small animal that began with the letter M. It was explained that sometimes the letter would appear before the adjective, and sometimes afterward.

The *S* sat in front of a screen in which was a window covered by half-silvered glass. An index card containing the restrictors was placed in a dark enclosure behind the window and was revealed by illuminating the enclosure. The *S* spoke his response into a microphone placed in front of him.

A trial consisted of the following events: (*a*) *E* spoke the name of the category; (*b*) 2 sec. later the first restrictor was illuminated; (*c*) after a .5-sec interval, the second restrictor was illuminated, and simultaneously an electric timer with a dc clutch was started; and (*d*) *S*'s verbal response activated a voice key that stopped the clock and terminated the trial. If *S* did not respond in 10 sec., the trial was terminated by *E*. Twenty warm-up trials preceded the experimental trials.

Results

Only correct responses (96%) are included in the following analyses. Responses were grouped into two classifications depending on whether the adjective came before or after the letter. Median latencies were obtained for each *S*'s responses in each of the two classifications. For each classification, group mean latencies were obtained by averaging medians for individual *S*s. These computations yielded a value of 1.00 sec. when the adjective came first, and a value of 1.45 sec. when the adjective came second. These means are significantly different, $t(15) = 5.612$, $p < .01$.

Two median RTs were also computed for each category, one for responses to adjective–letter stimuli and the other for responses to letter–adjective stimuli. For 43 out of 46 categories, *S*s produced a response more quickly if the adjective occurred before the letter restrictor; for only 3 categories, the reverse result occurred.

When the letter comes first, and *S* awaits the adjective, there is a great deal of uncertainty as to what the restrictor will be. An uncertainty interpretation correctly predicts that *S*s will take longer to respond in this condition.

General Discussion

The present experiments required *S*s to retrieve well-learned category members from semantic memory. In the first experiment, noun categories were paired with either an adjective or a letter, and these stimuli were presented in a randomly intermixed sequence of trials or in a nonmixed sequence. Reducing the uncertainty as to the identity of the forthcoming restrictor with advance cues did not facilitate RT, as an uncertainty interpretation would predict.

In the second experiment, noun categories were paired with both an adjective and a letter. An uncertainty interpretation seems to correctly predict that *S*s will take longer to respond when the letter restrictor is given first. In spite of the fact that an uncertainty interpretation makes a correct prediction in Experiment II, it in no way explains the results of Experiment I. What is needed is a model that can handle both of these results.

The Dictionary–Network Model

The model we propose attaches some importance to the finding of Experiment I that

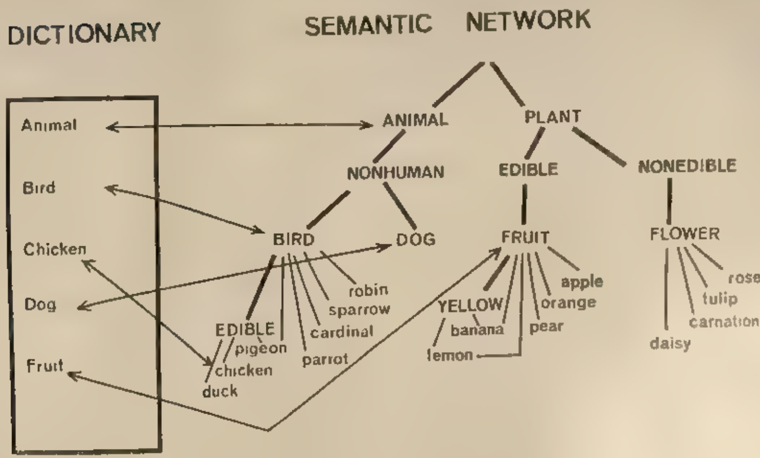


FIGURE 2. The dictionary–network model of semantic memory

the time to respond to an adjective stimulus is independent of whether *S* knew he would be seeing an adjective, but the time to respond to a letter stimulus is not independent. This finding suggests to us that intrinsically different processing strategies may be operating, and this notion is embodied in the following model.

First, we assume that when a word is stored in semantic memory, at least three kinds of information about that word are included: semantic, phonemic, and orthographic information. The assumption is reasonable since we know that *S*s use these three kinds of information when they search for particular words (Brown & McNeill, 1966).

We call the model the dictionary–network model, for reasons that will become obvious. It includes a "dictionary," as diagrammed in Figure 2, which is simply a listing of the lexical items of the language. As to the question of what constitutes a lexical item, we realize that it could be a word, a morpheme, an idiom, or some other unit, but we do not attempt to decide on the question of how the listing is ordered. The dictionary includes both phonemic and orthographic information about each of its entries, and, furthermore, each entry serves as an address telling where semantic information can be found. Thus, we have a content-addressable system in which each entry provides direct access to the portion of semantic memory that represents the meaning of that entry.

The semantic network is a highly complex arrangement of interconnected concepts or categories. Information about the subsets of each category as well as its supersets is included. Within each category [illegible] clusters of items that have [illegible] mon may exist. So, for ex[illegible] and yellow fruits may b[illegible] another in memory.

To illustrate, the word *animal* [illegible] an entry in the dictionary. Since [illegible] nary itself contains phonemic and [illegible] information about the word *ani*[illegible] use this information to answer questions like "What letter does the word [illegible] begin with?" To answer a question about the meaning of the word, it is [illegible] part of the semanti[illegible] information. Figure [illegible] pr[illegible] network that might be e[illegible] someone's semantic m[illegible] can use the information [illegible] semantic network to answer que[illegible] animal?" or "What is the name of [illegible] animal?"

Using this as a general [illegible] begin to describe how *S* manag[illegible] an answer to the kinds of questions he is asked in our experiments. How does he find (*a*) a category member to which a particular adjective is applicable (e.g., edible birds); and (*b*) a category member beginning with a [illegible] letter (e.g., an animal beginning with *D*)

Consider case *a* in which *S* must name an edible bird. We now assume that the retrieval process has at least three major steps. In keeping with our earlier view (e.g., Freedman & Loftus, 1971), the first step consists of entering the appropriate category—*birds*. The next step consists of finding an appropriate cluster of items and picking a member of it. Once this member is found, the third step

involves looking up the "name" of the item by reentering the dictionary via a pathway from that member to its name. Thus, *S* finds that the edible bird he has picked is called a *chicken*. At first glance, the notion that one could find a member without knowing the name of that member may seem awkward. But, on second glance, there is no reason why a particular type of category member, or any other concept for that matter, must necessarily have available a single name for it. Clearly, there can be concepts without names.[3] For example, one can comprehend the concept of a "navigational instrument used in measuring angular distances, especially the altitude of sun, moon, and stars at sea," without the name *sextant* being accessible (Brown & McNeill, 1966). The dictionary-network model is a reasonable system that allows access of conceptual meaning via the lexical entry in the dictionary, as well as permitting use of the meaning as a direct guide to the lexical entry.

Consider case *b*, in which *S* must name an animal beginning with the letter *D*. Again, *S* begins by entering the *animal* category, which is accomplished by looking up *animal* in the dictionary and going directly to the portion of memory that contains information on the meaning of *animal*. The next step is somewhat different for questions of this type. Here we assume that *S* systematically proceeds along pathways leading from animal instances to the dictionary representations of the names of those instances. This quasi-parallel simultaneous search proceeds toward the dictionary from all possible members (though a less dominant member may start off more slowly or slightly later). The dictionary is used to determine whether a given *animal* instance is one that begins with the letter *D*, and the search terminates when one is found.

Effect of Advance Knowledge

An explanation in terms of the model of the differential effects on adjective and letter trials of advance knowledge about the type of restrictor is straightforward. On adjective trials, the first step of the process is to enter the category. The next step, which involves finding the appropriate cluster to enter, cannot begin until *S* knows what cluster he is to enter. The *S* essentially must wait until the adjective is presented before he can begin this next step. Thus, knowing that an adjective is forthcoming should be of no help; that is, RT to O–A stimuli should equal RT to M–A.

For letter stimuli, a different situation exists. Here, the first step of the process is entering the category. The next step is a quasi-parallel simultaneous search toward the dictionary. That is to say, some number of pathways leading from category instances to the dictionary representations of those instances is traced. This step can be started during the interval between the presentation of the category name and the restricting letter if *S* knows a letter is forthcoming. Thus, RT to O–L stimuli should be faster than RT to M–L stimuli.

Effect of Order of Presentation of Attribute and Name Restrictors

Turning now to Experiment II, the dictionary–network model can easily handle the main result, viz., that *S* is faster to produce a category member subject to two restrictions when the adjective restrictor is presented first. The reasoning is as follows. Whether the adjective comes first or second, the first step of the retrieval process is to enter the noun category. When the adjective does appear first, *S* can then enter an appropriate cluster of items. At this point, he systematically proceeds along pathways leading from instances that are within that cluster to the dictionary representation of the names of those instances. The dictionary is used to determine whether any of the instances begin with the appropriate letter, and the search terminates when one is found. As before, this quasi-parallel simultaneous search can be started during the interval before the restricting letter appears, since *S* knows a letter is coming.

When the letter appears first, however, the picture is somewhat complicated. The *S* may do one of two things. First, after entering the category, he may begin to trace pathways toward the dictionary, attempting to find a member that begins with the given letter. Once he finds the member, he must go back

[3] This point is not particularly original. For example: "Concepts need not have names [Collins & Quillian, 1972, p. 316]," and "Concepts ... need not correspond in a one-to-one fashion with words. In fact, most concepts have no simple natural language equivalent [Rumelhart, Lindsay, & Norman, 1972, p. 216]." As to the question of what form of representation the nonverbal aspect of a concept takes, we do not attempt to decide that issue here.

to the semantic network to see if the adjective applies to it. This strategy requires an extra "trip" between the dictionary and the network. Alternatively, to save the trouble of this extra trip, *S* might simply wait until the adjective arrives, essentially converting a "letter-adjective" trial into an "adjective-letter" trial. Either way, the prediction is a slower RT when the letter appears first.

We might mention that our prediction for Experiment II is not obvious: Other quite reasonable models would not make this prediction. For example, one type of set-theoretic model assumes that once a category is entered, its members are sequentially searched in a relatively fixed order and that the search is terminated when an appropriate instance is located (see, e.g., Rosch, 1973). To the extent that retrieval involves such a successive consultation of the memory store, the speed of retrieval should not be a function of the order of presentation of adjective and letter. If each instance is successively examined to see if it meets both specifications, and the search terminates when such an instance is found, there is no reason to expect that presentation order will affect RT.

A slightly different serial model assumes that *S*s enter a subcategory defined by the category name and the first restrictor and begin searching in a fixed order for an instance which satisfies the second restriction. This model would predict that RT is faster when the letter is presented first. The reasoning is as follows: When the letter is presented first, a very small subcategory is usually defined, and finding an instance that fits the adjective restrictor should be a very easy matter. For example, if given *Animal-M-small*, searching the category *animals beginning with M* should result in rapid production of a small animal beginning with *M*, since roughly half of the members of that subcategory could conceivably (depending on *S*'s criterion) be considered small. When the adjective is presented first, the subcategory defined by the noun and the adjective is searched to find an instance that fits the letter restrictor. Since the standard of acceptability is much higher now (the instance must begin with a particular letter), the expected length of search for an instance meeting this standard is longer. In other words, a serial search for an instance meeting the letter restriction would inv[illegible]e consultation of many more items.

A Final Comment

Can anything be said abou[illegible] [illegible]e absolute speed of responding to a lette[illegible] [illegible]n adjective stimulus? Is one case more d[illegible] [illegible] than the other? At the present ti[illegible] [illegible]e have no definitive answer to this que[illegible] Response times depend to a large exte[illegible] character-istics of the pool of possible [illegible]ses; adjective and letter stimuli restri[illegible] to different response pools. If, for exa[illegible], the pool happens to contain a "dor[illegible]" category member, that is, one with a [illegible]quency of being given as an exempla[illegible] category, then (other things equal) th[illegible]nse will be relatively fast (Freedman [illegible]tus, 1971; Loftus & Suppes, 1972). In [illegible]ven experiment, then, whether adject[illegible]uli will be faster or slower than lette[illegible]uli depends largely on the particular adj[illegible] and letters we choose to include, and t[illegible] of responses to which they refer. F[illegible] reason, the dictionary–network model i[illegible]n the issue of the difficulty of adjecti[illegible] letter restrictors per se.

REFEREN[illegible]

BROWN, R., & MCNEILL, D. The "tip of the tongue" phenomenon. *Journ[illegible] of Verbal Learning and Verbal Behavior*, 196[illegible], 5, 325–337.

COLLINS, A. M., & QUILLIAN, M. R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

FREEDMAN, J. L., & LOFTUS, E. F. Retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 107–115.

GROBER, E. G., & LOFTUS, E. F. Semantic memory: Searching for attributes versus searching for names. *Memory and Cognition*, in press.

LOFTUS, E. F., & SUPPES, P. Structural variables that determine the speed of retrieving words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 770–777.

ROSCH, E. On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.

RUMELHART, D. E., LINDSAY, P. H., & NORMAN, D. A. A process model for long-term memory. In E. Tulving and W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

(Received November 17, 1973)

urnal of Experimental Psychology
974, Vol. 102, No. 6, 1123–1125

SHORT REPORTS

PARTIAL CATEGORY CUING: THE ACCESSIBILITY OF CATEGORIES

RICHARD E. PARKER [1] AND LINDA WARREN

Williams College

In order to investigate the hypothesis that the retrieval of some retrieval cues reduces the probability of retrieving other retrieval cues, experimental *S*s were given half of the category names from a categorized word list at the time of recall, while control *S*s were given no category names at the time of recall. The results supported the hypothesis and were interpreted as supporting a hierarchical model of retrieval from the memory store.

Rundus (1973) offered a conception of the retrieval process which by its own nature tended to limit the accessibility of information in memory. He introduced a hierarchically organized model of the memory store with retrieval from retrieval cues (RQs) down to words being accomplished by means of a series of random draws. All draws were assumed to be with replacement. The probability of any particular word being accessed from a particular RQ was equal to the ratio of that word's strength of association with the RQ to the sum of the strengths of association with the RQ of all the words associated with the RQ. This is the ratio rule. It was also proposed that both the presentation and recall of a word increased the strength of association between the word and the appropriate RQ. This combination of hypotheses produces the position that the retrieval of one word increases its own probability of retrieval, while at the same time reducing the probability for the other words associated with the same RQ. This ultimately leads to the termination of recall.

Another implication of the model is that the presentation of list words as cues for retrieval, by increasing their own strength of association with an RQ, should reduce the probability of retrieval for other words. This is the well-known negative effect of giving list elements as cues for retrieval, first reported by Slamecka (1968). In order to test this prediction directly, Rundus (1973) varied the number of words from a category given to *S* as cues. According to the model, giving proportionally more list elements as cues for a category should cause proportionally fewer other words from that category to be recalled. His results supported this prediction.

Rundus (1973) also proposed that the retrieval of RQs obeyed this same set of rules, which is the point of interest of this paper. In his model, the experimental list served as the center for a series of random draws with the RQs associated with the list serving as the pool. As before, the probability of retrieval was dependent on the ratio rule, and the presentation and retrieval of an RQ strengthened its association with the list. Thus, according to Rundus's model the presentation or retrieval of one RQ should increase its probability of retrieval, while at the same time reducing the probability of retrieval of the other RQs. Therefore, for the recall of a categorized word list, if *S*s were given only some of the category names as RQs, then the retrieval of the other RQs should be impaired. The purpose of this experiment was to test this higher level hypothesis.

Method. There were 16 control *S*s and 16 experimental *S*s with conditions run in a random order. Each *S* was paid $1 for his services.

Two sets of 40 words, each set consisting of 20 categories, 2 words per category, were selected from the Battig and Montague (1969) norms such that none were proper nouns. The sets of words were so composed since an examination of Tulving and Pearlstone's (1966) data indicated that the maximal effect would be obtained by the use of a large number of small categories. These category names and members were recorded on magnetic tape in a blocked fashion, with the category name first, followed by the two category members. For the experimental condition, the names of the categories were typed on index cards. For the control condition, a set of 10 numbers was typed on index cards.

Each *S* was tested individually. All *S*s were fully advised concerning the procedure and the composition of the lists before the presentation of the first list. All *S*s were told that their object was to recall as many category members as possible and that the category names were presented solely for their benefit. In addition, experimental *S*s were told that their optimal strategy would be to first recall as many words as they could without using the index cards on which the category names were typed. After doing this, they could try to use the index cards to facilitate further retrieval. These instructions were designed to decrease any tendency to recall cued categories earlier than noncued categories, always a concern in studies of cued recall.

After the instructions, *S* was presented, by means of a tape recorder, one of the lists of 20 category

[1] Requests for reprints should be sent to Richard E. Parker, who is now at the Department of Psychology, University of California, San Diego, P. O. Box 109, La Jolla, California 92037.

TABLE 1

CATEGORY RECALL DATA EXPRESSED AS MEAN PER S BY LIST, CONDITION, AND CATEGORY TYPE

Type of category	List 1		List 2	
	Experimental	Control	Experimental	Control
Cued	8.8	5.1	8.4	3.7
Noncued	2.9	4.8	2.6	4.6
Total	11.7	9.9	11.1	8.3

names and 40 category members at a 2-sec/item rate. If he was in the experimental condition, S next read aloud, from index cards at a 2-sec/item rate, 10 of the category names. According to the model, this should increase the strengths of association to the list of the 10 category name RQs. After this, the index cards were placed face down beside S, and he was given 2.5 min. for written recall. In order to cue each of the categories equally often, half of the experimental Ss read the category names that had been presented in the odd serial positions, while the other half read those presented in the even serial positions. In both cases, these names were read in the same order in which they had been presented originally.

If S was in the control condition, the procedure was the same except that the cards containing the 10 numbers were substituted for the cards containing the 10 category names.

Thirty seconds after the end of the recall interval each S was presented the other list of 20 category names and 40 words. This was deemed desirable since Rundus's (1973) results indicated a rather small effect for the manipulation. All Ss remained in the same condition for both the first and second lists.

Results and discussion. In order to determine whether or not giving experimental Ss category names as RQs introduced a differential guessing strategy, the intrusion data, by condition and trial, were compiled. All values varied around 9% of the total correct recall, indicating no differential guessing strategy. The words per category recalled by the two groups were also calculated. They varied between 1.6 and 1.7 for both lists, with the control group tending to recall slightly more words per category. However, the difference was clearly not significant. For the experimental Ss, there was no difference in the number of words per category recalled in cued and noncued categories.

The data of primary concern are the category recall data contained in Table 1. In Table 1 and hereafter, List 1 refers to the first list studied by S, regardless of its content, while List 2 refers to the second list studied by S, also regardless of the words in the list. For the purpose of data analysis, each control S was yoked with an experimental S who had studied the two lists in the same order. This yoking was used solely to provide a distinction between "cued" and "noncued" categories for control Ss. A category was scored as recalled if either category member was recalled. The category rec[illegible] data for List 1 indicated that experimental S[illegible] called slightly more categories than control [illegible] $(1, 30) = 6.51$, $p < .05$, effect size $\omega^2 = .04$. T[illegible] type of category, whether cued or noncued, had highly significant effect, $F(1, 30) = 52.93$, $p < .001$, $\omega^2 = .32$; and the Category Type $\times$ Conditi[illegible] interaction was also highly significant, $F(1, 30) = 41.10$, $p < .001$, $\omega^2 = .25$. Thus, the data indicate that experimental Ss recalled more cued [illegible] than control Ss but recalled fewer noncued [illegible] than control Ss. This is as predicted by [illegible] (1973) model. As is apparent from Table 1 [illegible] data for List 2 were very similar to the data f[illegible] List 1. Thus, the category recall data suppor[illegible] proposition that the retrieval of some cate[illegible] reduces the probability of retrieving other catego[illegible] The same effects were present for the word re[illegible] data: Experimental Ss recalled 70% of the cued words as opposed to 22% of the noncued words; control Ss recalled about 38% of each type of word

Another possible explanation of these [illegible] be based on the retention intervals of cued a[illegible] cued words. Maybe the effect is due to a [illegible] retention interval for the cued words. If t[illegible] a shorter retention interval for the c[illegible] for the noncued words, then it would not be possib[illegible] to evaluate the validity of Rundus's (19[illegible]) model. To test this possible explanation, experim[illegible] response sheets were divided into fifths on [illegible] of total number of correct words recalled. [illegible] data for both List 1 and List 2 are given in Tab[illegible] in terms of the percentage of total cued-wor[illegible] noncued-word) recall and in terms of t[illegible] amount of recall for each fifth. For each S wh[illegible] recalled some cued word [illegible]nd some non[illegible] ($n = 14$ for List 1, n [illegible] for List 2), the median recall position occupied by cued and noncued words was also calculated. For List 1, the mean of these median positions for cued words was 10.5 and for noncued words, 6.5, by randomization test, $p <$ .0[illegible] Thus, for experimental Ss the cu[illegible] appeared later in the recall outpu[illegible] words. This is the exact opposite of what a reten[illegible] tion interval argument would require. As the m[illegible]

TABLE 2

PERCENTAGE (AND MEAN NUMBER) OF WORDS RECALLED [illegible] CUED AND NONCUED CATEGORIES FOR EACH FIFTH OF THE EXPERIMENTAL Ss' RE[illegible]

Fifth	List 1		List 2	
	Cued	Noncued	Cued	Noncued
1	15.3 (2.14)	34.3 (1.59)	20.0 (2.75)	20.0 (.86)
2	18.6 (2.60)	24.3 (1.13)	19.8 (2.73)	20.6 (.89)
3	21.5 (3.01)	15.4 (.71)	19.7 (2.71)	20.9 (.90)
4	23.6 (3.31)	8.9 (.41)	19.7 (2.71)	20.9 (.90)
5	21.0 (2.94)	17.0 (.79)	20.7 (2.85)	17.7 (.76)
Total	100 (14.0)	100 (4.6)	100 (13.8)	100 (4.3)

amount of recall data in Table 2 indicate, this difference was basically due to the difference in total amount of recall. Since *S* knew that the category names would be available later, he may have tended to withhold cued-word recall. This may have contributed to this result. Any such tendency, of course, could have only had the effect of reducing the magnitude of the obtained results. In any case, the retention interval for cued words was not shorter than the retention interval for noncued words. Similarly, for List 2 the mean median word position for both cued and noncued words was 9.9. Thus, these data indicate that the argument of different retention intervals is not tenable. The lengths of the retention intervals did not differ in the required direction.

It should be noted that both the reading of the 10 category names and the freedom to use them later as cues were necessary in order to eliminate the above possibility of different retention intervals. Having *S* merely read the 10 category names and not to have them available later would create the problem of different retention intervals. It would be to *S*'s advantage to recall the cued categories first before he forgot their names. Furthermore, according to the model it was necessary to have *S* read the cues in order to establish different strengths among the RQs. Thus, the dual manipulation was necessary.

A second alternative explanation for the obtained results is that experimental *S*s spent more time trying to retrieve from the cued categories to the neglect of the noncued categories. However, this explanation does not seem viable for the current experiment both because of the long recall interval and because experimental *S*s recalled more cued words than noncued words in each fifth of their output. This is apparent from the mean amount of recall data in Table 2.

In conclusion, the results clearly indicate that when an *S* is given some of the category names from a categorized word list as RQs at the time of recall, his recall of these categories is enhanced. This result is consistent with the findings of Tulving and Pearlstone (1966). Moreover, at the same time, *S*'s recall of the other categories is reduced. This is the more surprising and revealing result. It is also consistent with Rundus's (1973) model of the recall process. While the present data support Rundus's model, it should be noted that his model is not entirely adequate. In particular, it is not able to account for sequential aspects of recall. Even in free-recall experiments the output order does become stereotyped (e.g., Bousfield, Puff, & Cowan, 1964). This is in addition to other designs in which *S* is specifically asked to order his output. In any case, the present data do support Rundus's conception of the recall process. By so doing, they support a hierarchical model of the memory store, with retrieval of one RQ reducing the probability of retrieving other RQs.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80** (3, Pt. 2).

Bousfield, W. A., Puff, C. R., & Cowan, T. M. The development of constancies in sequential organization during repeated free recall. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 489–495.

Rundus, D. Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 43–50.

Slamecka, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology*, 1968, **76**, 504–513.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

(Received August 3, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1126–1129

EVIDENCE FOR RETROACTIVE INTERFERENCE IN RECOGNITION FROM REACTION TIME[1]

ANDRIES F. SANDERS[2]
Institute for Perception, TNO, Soesterberg, The Netherlands
LESLIE WHITAKER AND CHARLES N. COFER
Pennsylvania State University

Retroactive inhibition (RI) was measured in the A–B, A–D; A–B, C–D; and rest-control paradigms. No RI was evidenced for any group by the conventional measure of number of correct responses in a recognition test of retention. However, strong indications of interference were found using a reaction time measure of response latency to the first-list pairs tested at retention. This extends the Postman and Stark results and indicates that the associative matching test is not entirely free of interference in the sense of generalized response competition.

The hypothesis that retroactive inhibition (RI) in the A–B, A–D paradigm may be due to suppression of the repertoire of first-list responses was proposed by Postman and his associates (Postman & Stark, 1969; Postman, Stark, & Fraser, 1968) on the basis of the fact that when learning of both lists and the test for retention of the first list are carried out by an associative-matching procedure no or very little RI occurs. This finding, of course, contrasts with that seen in studies in which the responses must be recalled. Consequently, Postman et al. (1968) suggested that RI might occur on the basis that first-list responses were not available at time of test, maintaining, however, that the first-list associations are still intact as they can be manifested on the matching test. It should be clear, however, that suppression of the first-list response repertoire is simply a mechanism additional to any specific item interference that may arise, as it does, for example, in the A–Br paradigm (Postman & Stark, 1969) and that item-specific interference cannot be ruled out because suppression of the first-list response set may occur (Postman & Underwood, 1973, p. 24).

Various experiments have been reported in which item-specific interference has been obtained, either in addition to or in the absence of evidence for loss of the first-list response set (Birnbaum, 1970, 1972; Delprato, 1972; Weaver, Duncan, & Bird, 1972; Weaver, Rose, & Campbell, 1971; Wichawut & Martin, 1971). Postman and Underwood (1973) have suggested limitations on the permissible conclusions to be reached from some of these studies, all of which involved the recall method. It has seemed to the authors that a further exploration of the associative-matching procedure itself would have merit. First, it would be desirable to replicate the findings of Postman and Stark (1969). Second, it is possible, with a matching test, to add information to what is provided by the number of correct matches by measuring reaction times. T[illegible] measure has often been proposed as an index of a[illegible] strength, although it may perhaps be give[illegible] tive interpretations in a matching test as op[illegible] to a recall test. In the matching results report[illegible] by Postman and Stark performance w[illegible] close to perfect in the A–D and certain other p[illegible], and one could therefore argue that perfo[illegible] the matching test was not sensitive to diff[illegible] associative strength but that reaction time [illegible] be. While it would obviously be possible t[illegible] an experiment in which matching performance would not be at a virtually perfect level, perfor[illegible] in the Postman and Stark experiment was very high. Since the conclusions concerning the supp[illegible] the first-list response pool come fro[illegible] ment, evidence from another me[illegible]sure of [illegible] ance seems desirable.

METHOD

The experiment involved three independent gr[illegible]ps of 10 female *S*s each, one for each of th[illegible] A–B, A–D; A–B, C–D; and A–B, control [illegible] paired associates of each list co[illegible] single letters as stimuli and one-syllable fo[illegible]-letter adjectives as responses.[3] A study–test procedure was followed, with the test part of each trial being in a recognition format, as described below. The control group performed an arithmetic task f[illegible] interval after A–B learning of the same duration as required for interpolated learning (IL) in the other groups.

Subjects. The *S*s were volunteers from classes in introductory psychology at the Pennsylvania State University who received credit toward their grades for participation in experiments. Assignment of *S*s to conditions was made randomly prior to arrival at the laboratory.

Materials. There were two sets of stimulus terms (A_1, A_2) which were combined to generate four basic

---

[1] This work was done during the period of A. F. Sanders' tenure as adjunct professor of psychology at the Pennsylvania State University. The help of William G. Hughes is gratefully acknowledged.

[2] Requests for reprints should be sent to Andries F. Sanders, Institute for Perception, TNO, Soesterberg, The Netherlands.

[3] The materials were the same as those used by Postman and Stark (1969). We are indebted to L. Postman for providing them.

lists. In each group *S*s were assigned so that four *S*s learned A_1–B_1 as original learning (OL), while two *S*s learned each of A_1–B_2, A_2–B_1, and A_2–B_2 as OL.

Several rules were employed in the construction of the pairs: (*a*) Stimulus letter and first letter of the response were different; (*b*) there was no phonemic overlap between the name of the stimulus letter and the response term; and (*c*) there was no apparent associative connection between the stimulus letter and the response. In the A–D list the responses to the same stimuli did not begin with the same letter as in the A–B list. Rule *c* was also applied in the construction of the A–D and C–D lists.

Procedure. A given study–test trial consisted, first, of the presentation of the 10 pairs one after another for 3 sec. each and, second, of 10 recognition tests in each of which one stimulus term was presented simultaneously with all 10 response terms. There was an interval of 2 sec. between the study trial and the first recognition test of the following test trial. An *S*'s response on the recognition test items advanced the memory drum so that the next stimulus term and the 10 responses were exposed. The interval between the tenth recognition test of a test trial and the next study trial was 2 sec. This procedure of alternate study and test trials continued until *S* had achieved 3 successive perfect trials following the first trial on which he responded correctly to all stimuli. Consequently, the first list was overlearned. There were 10 study–test trials of the second list (A–D or C–D), the first presentation of this list following 2 min. after the completion of OL. In the control group, this 2-min. interval was followed by 9 min. of work on an arithmetic task which required simple addition of digits, with the total number of digits added in successive 30-sec. periods being displayed to *S* as knowledge of results. At the completion of the interpolated task, there was a 3-min. rest interval after which *S* was given 3 consecutive test trials, each conducted as in OL, for the A–B list. (No study trials or knowledge of results were given on these test trials.)

In the study trials four different orders were used for the presentation of the pairs. The orders of presentation of the stimuli in the test trials were different from any of these, and there were 10 different orders of the responses in each test trial. In these orders each of the 10 response positions was correct once, with a different correct position (randomly chosen with restrictions) over the series of test trials. After each test trial *S* was told the number correct she had achieved on that trial. After the interpolated task, the first retention trial was given in the same trial order as the last test trial in OL.

Apparatus and manner of response. Three Stowe memory drums were used to make unnecessary any changing of tapes between lists. A reaction-time apparatus provided 10 buttons for *S*, and *S* was to press the one which corresponded to that one of the 10 responses simultaneously displayed on the memory drum which she thought to be correct. She was told to be sure of the response before making it but to respond as quickly as possible once she had made her choice. Thus, each *S* was urged to give fast but accurate responses. The numbers from 1 to 10 were affixed to the memory drum, each corresponding to one of the 10 responses displayed on each one of the 10 recognition tests of a test trial, and *S* called out the number of the response as she pressed the corresponding button. This enabled *E* to determine whether the response was correct. The *E* read the reaction time of the response from a digital counter in hundredths of a second. The *S*'s reaction, as indicated, advanced the drum, there being a ½-sec. delay between the button press and the appearance of the next stimulus and a new order of the 10 responses. Six Hunter timers were employed to control the drum during the study trial.

Results

First-list learning. The mean number of trials to the first perfect trial was 6.9 ($s^2 = 3.2$) for the A–D group, 7.3 ($s^2 = 4.7$) for the C–D group, and 8.7 ($s^2 = 9.1$) for the control group. Statistical analysis of these means yielded no significant differences.

To compute reaction times (RTs), the data for each *S* were handled in the following way. From the 10 RTs for each test trial the median and the first and third quartile (Q) values were determined. The median was used rather than the mean in order to avoid undue influence of extreme scores. Then, for the group of *S*s, the mean values of the median, Q_1, and Q_3 were determined over the values for individual *S*s on each trial. It is interesting to note that the mean of the median individual RT scores dropped significantly from the first trial on which all responses were correct to the final overlearning trial, $F\ (1, 27) = 24.4$, $p < .001$. This was true for all groups as shown by the failure of the Groups × Trials interaction to reach significance. The groups did not differ in the means of their Q_1, median, or Q_3 RT values on the final overlearning trial, $F\ (2, 27) = .220, .863, .905$, respectively.

Second-list learning. Although 10 trials were given on the second list, the first perfect trial was actually reached by Group A–D in 2.4 trials and by Group C–D in 2.3 trials. By this measure, Group A–D shows no negative transfer in comparison with Group C–D. However, the means (in sec.) of the median RT on the first perfect second-list trial were 4.26 for Group A–D and 3.25 for Group C–D. This difference is significant, $t\ (18) = 1.80$, $p < .05$. When difference scores were computed, i.e., when the mean of the median RTs for the final overlearning trial was subtracted from the mean of the median RTs for the first perfect IL trial, Group A–D had a significantly higher difference score than Group C–D, $t\ (18) = 2.31$, $p < .05$. These reaction time differences suggest the occurrence of negative transfer in Group A–D relative to Group C–D, although trials to criterion on the second list did not.

First-list retention tests. The three groups gave virtually identical and almost perfect scores for

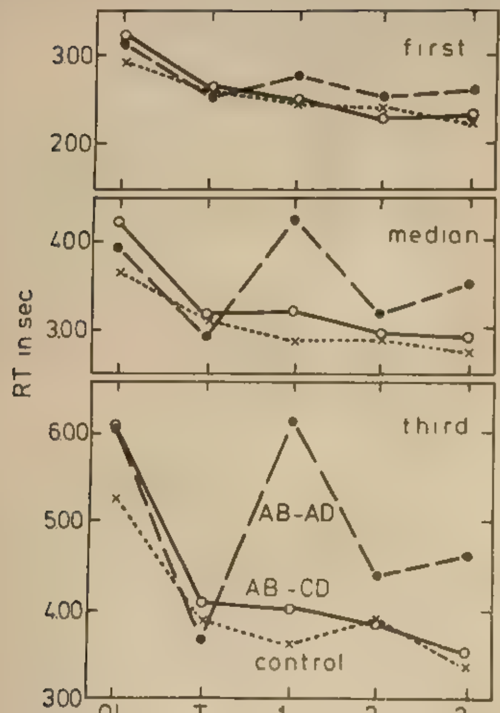


FIGURE 1. Means of quartile1 (Q_1), median, and Q_3 reaction times (RTs) for Groups A–D, C–D, and control on the first perfect trial in original learning (OL), the final overlearning trial in first-list learning (T), and the three retention test trials

correct recognition, thus confirming the results for this measure reported by Postman and Stark (1969). However, a different picture appears for the reaction time measure, as shown in Figure 1 for mean Q_1, mean median, and mean Q_3 scores for each of the three retention tests, the first perfect trial in OL, and the final overlearning trial (T). It may be seen that for all three measures the groups are comparable for T and for the first perfect trial in OL. In addition, they are quite similar for Q_1 over the test trials, the analysis of variance producing insignificant *F*s (2, 27) of 1.268, .641, and .882. However, the groups differ for the mean of the medians at Retention Tests 1 and 3, F (2, 27) = 10.304, $p <$ .01, and =3.825, $p < .05$; the similar smaller difference in the second test is not significant, F (2, 27) = .886. A similar pattern occurs for Q_3. Here the differences for Tests 1 and 3 are also significant, that for Test 2 not. The respective *F*s (2, 27) are 13.359 ($p < .01$), .888, and 4.512 ($p < .01$). Figure 1 shows that the significant differences arise from the longer reaction times for Group A–D, as compared with the other groups which do not differ from one another.

The foregoing analysis indicates that the fastest responses (those at Q_1) in the retest of A–D do not differ from those found in C–D or in the rest control. However, the responses with reaction times at the median and at Q_3 in A–D do differ from those in the other groups, significantly so at the first and third retention tests. On the first of these tests, it is worth noting that half of the *S*s in A–D show longer median RTs than the slowest *S* in C–D, and eight A–D *S*s show longer RTs than rest-control *S*s. There is no comparable dispersion of RTs for Group A–D in the final overlearning trial of OL.

Three other analyses of the data were made. The first was based on the notion that in the retention tests for A–B in Group A–D *S* may start by recalling the appropriate response from the second list, not making at once the switch from the second- to the first-list pool of responses. As he encounters the first-list response set over the tests of the first retention trial *S* might be able to reinstate the first-list response pool so that he would not make and have to reject the initial responses to later stimuli in the trial. Such a process would predict a decline in RTs over the first retention trial. An analysis was therefore made by listing for each *S* his median RT on Recognition Tests 1 through 10 during the first retention trial and averaging across differences for Tests 1 and 3 are also significant, that trial. While an analysis of variance for the rest-control group's data for test position was significant, F (9, 162) = 3.48, $p < .01$, the analysis was not significant for the other groups. Hence, there is no evidence for recovery of the List 1 response pool over the 10 recognition tests in either Groups A–D or C–D.

The second analysis followed that of Runquist (1957) who related the number of correct responses to a given stimulus–response (S–R) pair with its resistance to RI in a recall task. In the present study, RTs on the first retest for pairs which yielded incorrect responses during OL were compared with RTs for pairs always correct. The analysis showed that the median RT for previously missed pairs was greater than the median RT for pairs always correct for 22 of 27 *S*s. (Three *S*s made no errors in acquisition.) This is significant by sign test at $p < .05$. The 5 *S*s whose median RT for missed responses was less than or equal to the median RT for their correct responses were distributed as follows: 2 controls and 3 AB–CD *S*s. It is difficult to argue that this is strong evidence for strength of specific S–R bonds, however, particularly in the light of a correlational analysis by *S* of the RTs on the following trials: (*a*) first OL 100% correct trial, (*b*) final OL trial, (*c*) first retention test, (*d*) second retention test, and (*e*) third retention test. While some correlations reached a significant level ($r = .540$, $t \geq 1.8$) for some *S*s, no correlations showed a consistent trend either for all *S*s or within each group. Thus, there was no evidence to support the role of individual associative strengths as measured by RT in interlist transfer or retention.

The third analysis was an attempt to determine whether the responses in Group A–D whose reaction times were at Q_1 were distinguished from the rest of the responses. The only factors that could be examined were the serial position of the item in the first retention trial and the position of the correct

button for each item on that trial. Neither an analysis of variance nor a correlational analysis yielded significant relations between either of these factors and reaction times of first-list retention-test responses.

Discussion

There are four major findings from this experiment, aside from the replication of Postman and Stark's (1969) failure to find RI in A–D on an associative-matching test so far as correct responses are concerned. (*a*) Reaction times in second-list learning were elongated for Group A–D with respect to those for C–D. (*b*) Reaction times at the first retention test for List 1 were elongated in Group A–D as compared to the other groups, and this elongation persisted over the remaining retention tests. (*c*) Some responses in A–D (those at Q_1) did not show this elongation. (*d*) Efforts to relate specific associative strength, position in the test order, serial position of the item, and position of the correct button failed to relate to reaction times on the first retention test.

If one defines negative transfer and RI as the occurrence of any deficiencies of performance in the A–D paradigm as compared to C–D and the rest control, then both have been found in this study in the reaction time measure, despite the absence of negative transfer in terms of trials to criterion in second-list learning and the absence of RI in terms of correct responses on the matching test. The interpretation of these findings is somewhat puzzling in view of the presence of the correct responses on each test in every trial.

One could interpret operationally defined RI in terms of diminished associative strength of first-list responses. Such an interpretation finds no direct support in our data, however. A second interpretation that fits both the negative transfer and the RI found is a response competition notion. Thus, negative transfer in terms of reaction time would be produced in the A–D case if, when *S* sees a stimulus term, he thinks of a first-list response. He will, of course, fail to find this response in the set of alternatives provided and will have to find in the set that response which goes with the stimulus. The same process would produce RI if the second-list responses, at the end of 1L, now are dominant and one occurs when a stimulus is presented. Again *S* would not find that response among the set of alternatives on the first-list retention test and would either have to identify the correct response in the set or attempt to remember what it was for that stimulus and then confirm the memory by checking for the response among the alternatives.

This kind of interpretation, which at least seems plausible, essentially amounts to that of generalized response competition. Hence, the negative transfer and the RI found in the present data suggest the presence of generalized response competition in the associative-matching situation, not detected by the simple measure of number of correct choices.

References

Birnbaum, I. M. Response selection and retroactive inhibition. *Journal of Experimental Psychology*, 1970, **85**, 406–410.

Birnbaum, I. M. General and specific components of retroactive inhibition in the A-B, A-C paradigm. *Journal of Experimental Psychology*, 1972, **93**, 188–192.

Delprato, D. J. Pair-specific effects in retroactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 566–572.

Postman, L., & Stark, K. Role of response availability in transfer and interference. *Journal of Experimental Psychology*, 1969, **79**, 168–177.

Postman, L., Stark, K., & Fraser, J. Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 672–694.

Postman, L., & Underwood, B. J. Critical issues in interference theory. *Memory & Cognition*, 1973, **1**, 19–40.

Runquist, W. N. Retention of verbal associates as a function of strength. *Journal of Experimental Psychology*, 1957, **54**, 369–375.

Weaver, G. E., Duncan, E. M., & Bird, C. P. Cue-specific retroactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 362–366.

Weaver, G. E., Rose, R. G., & Campbell, N. R. Item-specific retroactive inhibition in mixed-list comparisons of the A-B, A-C and A-B, D-C paradigms. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 488–498.

Wichawut, C., & Martin, E. Independence of A-B and A-C associations in retroaction. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 316–321.

(Received August 16, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1130-1132

CLARIFICATION OF THE ROLES OF ABSOLUTE AND RELATIVE FREQUENCY ON LIST DIFFERENTIATION

MILDRED MASON[1] AND LEONARD KATZ
University of Connecticut

Three different ratios (.20, .50, and 1.00) of List 1 to List 2 trials were investigated under two levels of absolute frequency (6 or 12 trials apportioned between the two lists). A list differentiation test immediately following List 2 learning employed three response categories ("List 1," "List 2," "neither list"). Results clearly pointed to the effects of absolute frequency on list differentiation. Contrary to Winograd's finding, list differentiation improved consistently as the relative frequency of List 1 to List 2 trials increased.

The notion of list differentiation (*S*'s discrimination of the previous list membership of test stimulus items) has played a large role in interference theories of memory (see Keppel, 1968). Until relatively recently, however, the phenomenon has been somewhat lacking in both data and explanation. Winograd (1968) found that given an equal number of trials on both lists, list differentiation (LD) improved with absolute frequency. An explanation of LD has been complicated by the further finding (Winograd, 1968; Winograd & Smith, 1966) that an unequal number of learning trials on two lists produces better LD than do balanced frequencies.

However, the issue of whether unequal frequencies do, in fact, improve LD has not been satisfactorily settled. Reported relative frequency effects may be questioned on two counts. First, as Winograd (1968) pointed out, absolute frequency and relative frequency are unavoidably confounded in designs where the number of trials on one list are held constant while varying the number of trials on the second list (e.g., in comparing a three trial–three trial group with a three trial–six trial group). Second, relative frequency effects may possibly be due to differential guessing strategies available to *S*s in the unbalanced frequency conditions. In previous studies LD was tested following List 2 learning by assuring *S*s that every item seen on the LD test had either been on List 1 or on List 2, with *S*s required to make a forced choice into one of the two categories. Conceivably, *S*s with many trials on one list and few trials on the other list could have assigned unfamiliar test words to the list with the fewer trials. This would artificially inflate correct responses on the weaker list and would account for the putative superiority of the three trial–one trial group to the three trial–three trial group.

Bower (1972) also has raised the possibility of differential guessing strategies in an attempt to reconcile the Winograd data on unequal frequencies with the Anderson and Bower (1972) list marker theory of list differentiation. The data on the superiority of unbalanced frequencies on LD constitute an embarrassment to a strength theory of LD as well as to the list marker theory.

The present study was designed to clarify the role relative frequency plays in LD. Two major methodological differences exist b[illegible]he present study and previous studies. First, the independent variables of absolute and relative frequency have been unconfounded by varying the distribution of List 1 and List 2 trials within fixed levels of absolute frequency. This technique has the further advantage of keeping the total temporal dispersion equal for all relative frequencies within any one level of absolute frequency (see Hintzman & Waters, 1970). Second, the list differentiation test used in the present study presented *S* with a forced choice into three "list" categories, instead of two: *S*s were told that each word on the LD test had either been on the first list they learned, on the second list they learned, or not on either list.

In all other important aspe[illegible] position, duration of word exposure, intertrial interval, and interlist interval) the present study attempted to replicate or remain similar to the Winograd (1968) studies.

Method. The desig[illegible] the total number of trials for List 1 and List 2 combined was one of two levels of absolute frequency: a total of 6 or a total of 12 trials. Within each level, three relative frequencies of List 1 to List 2 trials were employed; the ratios were .20, .50, and 1.00. Thus, *S*s in the 6-trial condition received one of the following conditions: 1 trial on List 1, 5 trials on List 2 (1–5); 2 trials on List 1, 4 trials on List 2 (2–4), or 3 trials on each list (3–3). Counterparts in the 12-trial condition were 2–10, 4–8, and 6–6.

Three lists of 20 nouns each were matched for frequency on the Thorndike–Lorge (1944) count; 48 of the 60 nouns were taken from the lists used by Winograd (1968). Each *S* received two of the lists for learning, with the third list serving as the set of distractor items on the LD test. The three lists were counterbalanced across conditions, each list serving as first list, second list, and distractor list for ⅓ of the *S*s in each condition. Five different orders of each list were used. All words were projected onto a screen in the front of the room by a Kodak Carousel slide projector activated by Hunter

[1] Requests for reprints should be sent to Mildred Mason, Department of Psychology, University of Connecticut, Storrs, Connecticut 06268.

timers to advance at 4-sec. intervals. The intertrial interval was 15 sec.; the interlist interval was 45 sec.

The *Ss* were initially instructed that they were participating in an experiment on memory and that they were to try to learn each word as it appeared on the screen. At the end of List 1 learning, *Ss* were instructed that they would be shown a second list of words and that they were to learn those words also. After the last trial of List 2, test instructions were given and test materials handed out. The LD test itself started 35 sec. after the last word of List 2 was shown, and consisted of the 60 words from all three lists arranged in random order on three pages. The LD test required *Ss* to make a forced choice for each word into one of the three categories "List 1," "List 2," and "neither list" on the test booklet. The LD test was timed with a stopwatch, with *E* saying "next" at 4-sec. intervals indicating that *S* was to proceed to the next word. Two different orders of the LD test were used, with half of the *Ss* in each condition receiving each order.

The *Ss* were 72 undergraduates, who participated in partial fulfillment of requirements for an introductory psychology course. There were 12 *Ss* randomly assigned to each condition. Groups of between 1 and 4 *Ss* served simultaneously.

Results and discussion. Figure 1 presents the mean number of correct list identifications for List 1 and List 2 combined. List differentiation appears to be a function of both absolute and relative frequency. An analysis of variance was performed on the combined number of correct List 1 and List 2 identifications for each *S* in each level of absolute and relative frequency. The effect of absolute frequency was significant, $F(1, 66) = 15.65$, $p < .001$. The effect of relative frequency was in the *opposite* direction of that found by Winograd (1968). List differentiation was best when the two lists were presented an equal number of times and was worst when List 1 was presented only 20% as often as List 2. The linear component of the relative frequency effect was significant, $F(1, 66) = 5.22$, $p < .05$; the quadratic component was negligible.

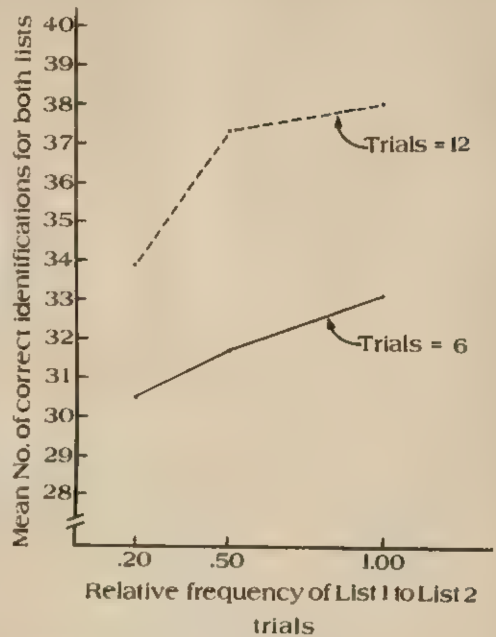


FIGURE 1. List differentiation as a function of absolute frequency (total trials List 1 plus List 2) and relative frequency of List 1 : List 2.

The mean square for the interaction was also nonsignificant and small.

The mean number of correct identifications in each of the three response categories for the six conditions is shown in Table 1. The "Selected" data columns present the mean number of times a particular category was selected, ignoring whether the choice was correct or incorrect. The contention that a forced-choice LD test into two categories biases LD in favor of the less frequent list is supported by the "Selected" data in Table 1. It is precisely at the most unbalanced distribution of trials (relative frequency = .20) that the "neither list"

TABLE 1

MEAN NUMBER OF CORRECT IDENTIFICATIONS WITHIN EACH RESPONSE CATEGORY AND MEAN NUMBER OF TIMES EACH RESPONSE CATEGORY WAS SELECTED

Relative frequency and condition	List 1		List 2		Neither	
	Correct	Selected	Correct	Selected	Correct	Selected
.20						
1–5	12.42	17.83	18.00	19.17	16.33	23.00
2–10	15.41	17.92	18.42	20.25	18.92	21.83
M	13.92	17.88	18.21	19.71	17.63	22.42
.50						
2–4	13.50	16.00	18.17	22.42	18.75	21.58
4–8	18.08	19.58	19.25	20.75	19.25	19.67
M	15.79	17.79	18.71	21.59	19.00	20.63
1.00						
3–3	15.67	17.83	17.42	20.08	19.58	22.08
6–6	18.83	19.58	19.25	20.33	19.92	20.08
M	17.25	18.71	18.34	20.21	19.75	21.08
Overall *M*	15.65	18.12	18.42	20.50	18.79	21.37
Hit rate	.86		.90		.88	

TABLE 2

Assignment of Errors within Response Categories

Absolute frequency and condition	"Neither" list errors		List 1 errors		List 2 errors	
	List 1	List 2	"Neither"	List 2	"Neither"	List 1
6 trials						
1–5	44	0	77	14	3	21
2–4	14	1	32	46	6	16
3–3	5	0	22	30	10	21
12 trials						
2–10	13	0	35	19	2	17
4–8	9	0	5	18	1	8
6–6	1	0	0	14	2	7

category is used more than would be predicted by chance; χ^2 (2) = 8.89, $p < .02$.

The manner in which the total number of errors within each response category was distributed within the other two response categories is shown in Table 2. The data clearly show more assignments of List 1 items to the distractor category ("neither list") in Conditions 1–5 and 2–10 than in the other condition; χ^2 (2) = 42.61, $p < .001$, for the 6-trial conditions, and χ^2 (2) = 25.11, $p < .001$, for the 12-trial conditions. If there had been no distractor category and if we assume that List 1 items assigned to the distractor category would have been placed in the List 1 category, it is easy to see from the data of Table 2 that there would be a higher frequency of "correct" List 1 responses in the most unbalanced frequencies—the result reported by Winograd (1968). Thus we may conclude that previous reports of LD being poorer in the balanced frequency conditions are indeed due to *S*s in these conditions not having the optimal guessing strategy available to *S*s in the unbalanced frequency conditions (i.e., "if list membership of an item is not known, assign it to the weaker list").

The finding that LD *improves* as relative frequency increases is difficult to account for with either a strength theory or a list marker theory. In addition, both theories run into other specific problems with the data. A strength theory of LD would predict that more List 2 responses would be assigned to List 1 as the number of trials on both lists approached equality. As can be seen from Table 2, this did not occur at the lower level of absolute frequency. At the higher level of absolute frequency, the number of List 2 errors assigned to List 1 was actually less for the 6–6 group than for the 2–10 group, χ^2 (1) = 4.17, $p < .05$. Similarly, strength theory would predict that more List 1 responses would be assigned to List 2 when relative frequency = 1.00 than when relative frequency = .20. This did occur in the lower absolute frequency condition, χ^2 (1) = 5.81, $p < .02$, but not in the higher absolute frequency condition, where the trend was rever[illegible]. We suggest that subjective organization of the lists accounts for these results and conclude with Winograd (1968) and Hintzman and Waters (1970) that absolute frequency affects list organization in addition to the effect it has on strength.

A list marker theory would predict a high[illegible] rate for the distractor items than for Li[illegible] as can be seen from Table 1, did not occ[illegible] importantly, a list marker theory would predict that distractor items have an equiprobable ch[illegible] of incorrectly acquiring a List 1 or a List 2 marker. With the counterbalanced stimulus lists used in this study, the possibilities of associative generalization are the same. The data from Table 2 clearly show that distractor items were incorrectly assigned to List 1 and not to List 2. Furthermore, distractor items are most likely to be [illegible]gned to List 1 when List 1 is weak relative to Li[illegible] 2, χ^2 (2) = 39.71, $p < .001$, for 6 trials, and χ^2 [illegible] 9.74, $p < .01$, for 12 trials. The confusion is cl[illegible] temporal dimension.

Hintzman and Waters (1970) showed th[illegible] LD performance was based on simultaneo[illegible] pendent discriminations on a frequency and a recency dimension. We would [illegible] or temporal, dimension to account for LD [illegible]ing as the number of trials on both lists reached e[illegible]ty. Within either level of absolute frequency [illegible] time is equal among levels of [illegible]ative fr[illegible] Results of the present experim[illegible] strongly sug[illegible] that the way this total te[illegible] dimension is partitioned into two successive time blocks ("List 1" and "List 2") affects temporal discriminability and hence LD. Whether the manner in which the total time is partitioned into two ca[illegible] the retention of temporal information or affects the decision process involved (Hinrichs, 1970) is a matter for future inves[illegible]

RE[illegible]

Anderson, J. R., & Bow[illegible] processes in free recall. P[illegible] 97–123.

Bower, G. H. Stimulus-s[illegible] bility. In A. W. Melton & E. [illegible] *in human memory*. Washingto[illegible]

Hinrichs, J. V. A two-proce[illegible] [illegible]eory for judgment of recency. *Psychological Review*, 1970, 77, 223–233.

Hintzman, D. L., & Waters, R. M. Recency and frequency as factors in list discrimination. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 218–2[illegible].

Keppel, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, [illegible]

Thorndike, E. L., & Lorge, I. *The tea[illegible] 30,000 words*. New York: Columbia Un[illegible] College, Bureau of Publications, 1944.

Winograd, E. List differentiation as a function of frequency and retention interval. *Journal of Experimental Psychology Monograph Supplement*, 1968, 76(2, Pt. 2).

Winograd, E., & Smith, W. S. List differentiation with varied trials on both lists. *Science*, 1966, 152, 1011–1012.

(Received August 23, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1133-1135

RECALL ACCURACY OF EIDETIKERS[1]

CHARLES J. FURST,[2] KENNETH FULD, AND MICHAEL PANCOE
Dartmouth College

A large group of second-, third-, and fourth-grade schoolchildren were shown pictures and asked to recall them verbally. The children were also asked to report whether or not they could see a projected image of the stimulus picture after it had been removed. Recall scores were compared for those children who reported such eidetic images and those children who did not. It was found that the eidetikers recalled considerably more material than noneidetikers, and this result was discussed in the context of questions raised about the nature of eidetic imagery.

In the recent revival of interest in eidetic imagery (EI), the question has been raised as to the uniqueness of this memory phenomenon. Neisser (1970) suggested that since there have been no demonstrable differences in capacity between children who report EIs and those who do not, EIs should be regarded as ordinary visual memories which are somewhat arbitrarily construed by their beholders to be located in front of their eyes. The most thorough available modern evidence fails to demonstrate a strong superiority in visual recall for eidetikers (Leask, Haber, & Haber, 1969), although Haber and Haber (1964) had earlier reported a slight difference in favor of these children.

We recently had the opportunity to check on the recall accuracy of eidetikers during the course of screening a large number of children for EI (for subsequent electroencephalographic studies), and these results are reported here. Free-recall reports of visually presented material were obtained and scored for accuracy, using an open-ended but reliable scoring system. In view of Neisser's (1970) assertion, it was decided that the most relevant test would be of the null hypothesis that no differences in recall accuracy would exist between children reporting EIs and those not reporting EIs. Accordingly, no distinction was made in the recall reports of the children between those items recalled from an EI and those recalled from other types of memory.

Method. A total of 192 children, comprising the second, third, and fourth grades of a local public school, were tested for eidetic imagery. Of these, 21 satisfied the criteria for being included in the eidetic group for this study. These criteria were a reported EI to both of the stimulus pictures, the duration of at least one of which was greater than 15 sec., and visual scanning (as judged by *E*) accompanying the reported eidetic episodes. The remaining children constituted the control group, with the exception of 7 children who were not included in the study because they reported EIs that did not meet the above criteria.

[1] This research was supported by Grant HD-07037 to the first author from the National Institute of Child Health and Human Development. The authors wish to thank the students and teachers of the White River Junction (Vermont) Elementary School for their cooperation.

[2] Requests for reprints should be sent to Charles J. Furst, Department of Psychology, Dartmouth College, Hanover, New Hampshire 03755.

The audiotapes of the screening sessions were incomplete or unscorable for 38 *S*s, including 3 from the eidetic group. This left an eidetic group of 18 *S*s and a control group of 129.

The children were individually conducted to a room near their classroom, where they were told that *E*s wished to show them pictures and ask them how they "see things and remember them." They were first given a demonstration of a chromatic afterimage to a colored square on a gray background. This was to aid them in distinguishing between "seeing" an object that was no longer present from merely remembering it.

Each *S* was then shown the first stimulus picture, a detailed colored illustration from a child's storybook, mounted on a gray cardboard backing. The mounted picture was placed on an easel in front of *S* for 30 sec. and then removed, revealing a blank cardboard surface on the easel of the same color as the picture backing. The child was asked to look at the board and to report whether or not he could still see any parts of the picture there. If *S* replied that he could, he was pressed to make sure that he distinguished this "seeing" from remembering and also as to the location of this image. He was asked to describe the image and to report when it had disappeared. When the image had disappeared, the child was asked to recall everything he could remember about the picture. All *S*s not reporting EIs were asked to report all details of the pictures which they could remember. The session was recorded on audiotape. In addition, *E* recorded the duration of any reported EIs and his judgment of whether or not they were accompanied by visual scanning of the easel board. At the conclusion of the recall of the first picture, a second stimulus picture was presented in a similar fashion.

The audiotapes were scored by two persons for number of items accurately reported. The system used gave 1 point for each object in the picture correctly reported and 1 point for each "attribute" of an object. Attributes were defined to include such things as color, article of apparel, or activity, but not location in the picture. For example, a report correctly stating that the picture contained *two elephants, wearing blue bow ties, standing in the road* would score 1 point each for *elephant*, *blue*, *bow tie*, and *road*. An additional point would be added

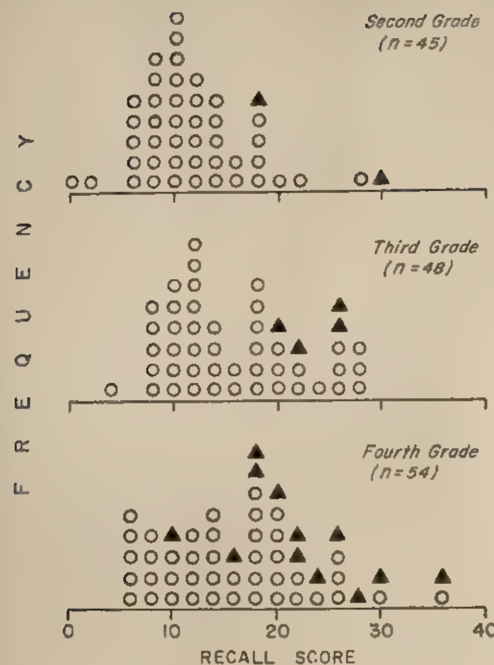


FIGURE 1. Frequency distributions of recall scores for controls (open circles) and eidetikers (triangles).

for stating that there were a pair of elephants but not for the verb *standing*, which would have been interpreted as indicating location rather than an activity. Interscorer reliability was quite high with this method ($r = .99$). No distinction was made between items reported from an EI and those reported from a subsequent memory after the EI had faded. Each S was given a recall score which was the sum of his scores for both stimulus pictures.

Results and discussion. Mean number of items correctly reported is shown in Table 1. An analysis of variance shows the difference between eidetikers and controls to be significant, $F\ (1, 141) = 73.1$, $p < .001$, as well as a main effect of school grade, $F\ (2, 141) = 34.9$, $p < .001$. The interaction was not significant ($F < 1$). The children reporting EIs were clearly superior to controls in their recall

TABLE 1

MEAN VISUAL RECALL SCORES FOR EIDETIC AND CONTROL GROUPS

Group	Grade: Second	Third	Fourth
Eidetikers	24.5	24.3	22.8
SD	7.8	3.2	6.9
Controls	11.9	16.1	16.3
SD	5.3	6.3	7.1

by a ratio of about 3:2. Distributions of recall sc[illegible] are given in Figure 1. An examination of Fi[illegible] 1 reveals that although the eidetikers were gene[illegible]ally near the top of the distributions, there were exceptions, and there were many noneidetic children wh[illegible] did as well as the eidetikers.

These results contradict Neisser's (1970) ass[illegible] that eidetikers do not have a demonstrably [illegible] recall of visual scenes than other people. H[illegible] Haber (1964) reported only slight superi[illegible] visual recall for their eidetikers, although th[illegible] scores were not given. Haber and Haber's [illegible] categories were predefined (i.e., *E*s decided be[illegible] hand what the admissible items would be), wh[illegible] the system used here was more open-ended [illegible] how a visual scene is coded tends to be a hi[illegible] idiosyncratic affair (Bartlett, 1932), an open-end[illegible] scoring procedure might be a more sensitive test [illegible]

The study reported here did not separate out information supplied from a supposed EI and information recalled subsequently through a more commonplace visual memory. As stated above, thi[illegible] was because it was desired to test the hypoth[illegible] differences between children who report E[illegible] others. We often found it diffi[illegible] to determin[illegible] when a child reported an item from an EI and when it was being supplied from memory.[3] R[illegible]son (1969) suggested that the presumed super[illegible] of an eidetiker might be due t[illegible] prolonge[illegible] time, since he would be able to [illegible]ew" th[illegible] some time after the original [illegible]cture had been [illegible]moved. Some such explana[illegible] [illegible]is would be the most direct. However [illegible] are still viable. For ex[illegible], it is pos[illegible]ible that children who report EIs are more talk[illegible] [illegible] hence, report more ite[illegible] during r[illegible] interestingly, it might [illegible] that children with the most vivid visual imagery of a perfectly [illegible]dinary type tend to localize their images in front of their eyes.

Although the eidetikers in this study showed a clear superiority in visual rec[illegible] was not so far off the distributions of sc[illegible] their classmates that one would be justified in con[illegible]cluding from these data alone that they posse[illegible]sed a special ability. Stromeyer and Psotka (1970) have recently reported a remarkable objective demonstra[illegible]tion that an eidetic image can be "photograph[illegible]" in the sense of being highly detailed in texture, at least for one talented adult S. It is still unresolved whether or not the type of eidetic imagery commonly reported by children is of this type. Stromeyer and his co-workers (C. F. Stromeyer, personal communication, 1971) have been unable to demonstrate this, but it remains possible that childhood eidetic

[3] Some of the children reported brief EIs which disappeared before material from it was verbally reported. Other children started out reporting material "seen" on the board and then afterward said that most of the report had been from memory, when the EI had faded. For the 20 reports from which scores for eidetically seen and remembered material could be distinguished, 45% of the material correctly reported came from the EI. There was great variability around this average percentage ($s = 31$), with no relationship between the total recall score and the percent of the material recalled from the EI.

imagery is extremely detailed but fragmentary, and therefore, different from ordinary visual memory.

REFERENCES

BARTLETT, F. C. *Remembering.* Cambridge: Cambridge University Press, 1932.
HABER, R. N., & HABER, R. B. Eidetic imagery: I. Frequency. *Perceptual and Motor Skills,* 1964, **19**, 131-138.
LEASK, J., HABER, R. N., & HABER, R. B. Eidetic imagery in children: II. Longitudinal and experimental results. *Psychonomic Monograph Supplements,* 1969, **3**(Whole No. 35), 25-48.
NEISSER, U. Visual imagery as process and as experience. In J. S. Antrobus (Ed.), *Cognition and affect.* Boston: Little, Brown, 1970.
RICHARDSON, A. *Mental imagery.* New York: Springer, 1969.
STROMEYER, C. F., & PSOTKA, J. The detailed texture of eidetic images. *Nature,* 1970, **225**, 346-349.

(Received September 4, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1135-1137

PROCESSING THE TERMS *RIGHT* AND *LEFT*: A NOTE ON LEFT-HANDERS[1]

GARY M. OLSON[2]
Michigan State University

KEVIN LAXAR
Naval Submarine Medical Research Laboratory, Groton, Connecticut

A simple word–picture verification task using the terms *right* and *left* was run with left-handed *S*s. Unlike right-handers, these *S*s showed no evidence for processing asymmetries between *right* and *left*. In all other respects, however, their performance was comparable to right-handers.

Olson and Laxar (1973) reported a series of investigations which supported the claim that the mental representation of *right* is less complex than that of *left*. Their experiments eliminated simple scanning biases as the source of their effects and suggested rather that asymmetries in the reaction times to simple word–picture displays were due to characteristics of central information processing. These experiments were all run with right-handed *S*s, and it would be of interest to know what comparable performance is like with left-handers. The present paper provides evidence on this question.

The assumption that *right* is the less complex term is quite plausible for right-handers. Their handedness coincides with cultural and social conventions to make right the natural reference direction in the sagittal plane, just as one can argue that aboveness and forwardness serve as reference directions in other planes (Clark, 1973; Olson & Laxar, 1973). However, the picture is much less clear for left-handers. Three possibilities seem to have equal a priori plausibility: (*a*) because left-handers live in an essentially right-handed world, their internal model of space conforms to that of right-handers; (*b*) although social and cultural conventions favor right as the reference direction, the left-hander's handedness is the more dominant factor, yielding *left* as the simpler term; and (*c*) since the criteria leading to *a* and *b* are in conflict, no performance asymmetry comparable to that found with right-handers should emerge. This last possibility could signify either the absence of an underlying conceptual asymmetry or conflict between alternative sets of representations.

Method. The design and logic of this experiment were identical to Experiment I in Olson and Laxar (1973). There were four different displays, containing either the word RIGHT or LEFT in block letters centered in a small square and a black dot located either to the left or right of the square. The *S*s' task was to decide as quickly as possible whether the word correctly named the side on which the dot appeared. The displays in which the word correctly named the side were the *true* conditions, and the other two displays were the *false* conditions.

The displays were presented to *S*s in blocks of 28 trials, each display appearing seven times per block. The first 4 trials in a block were warm-ups and were not analyzed. Each *S* was given four blocks of trials for a total of 96 experimental trials plus 16 warm-ups, all within a single 45-min. session. The order of presentation within each block was independently randomized for each *S* and each block.

[1] This research was carried out under the auspices of the Bureau of Medicine and Surgery, Navy Department, Research Work Unit MF51.524.004-2002DX5G. The opinions and assertions contained in this paper are the private ones of the authors and should not be construed as official or as reflecting the views of the Navy Department or the Naval Submarine Medical Research Laboratory.
[2] Requests for reprints should be sent to Gary M. Olson, Department of Psychology, Michigan State University, East Lansing, Michigan 48824.

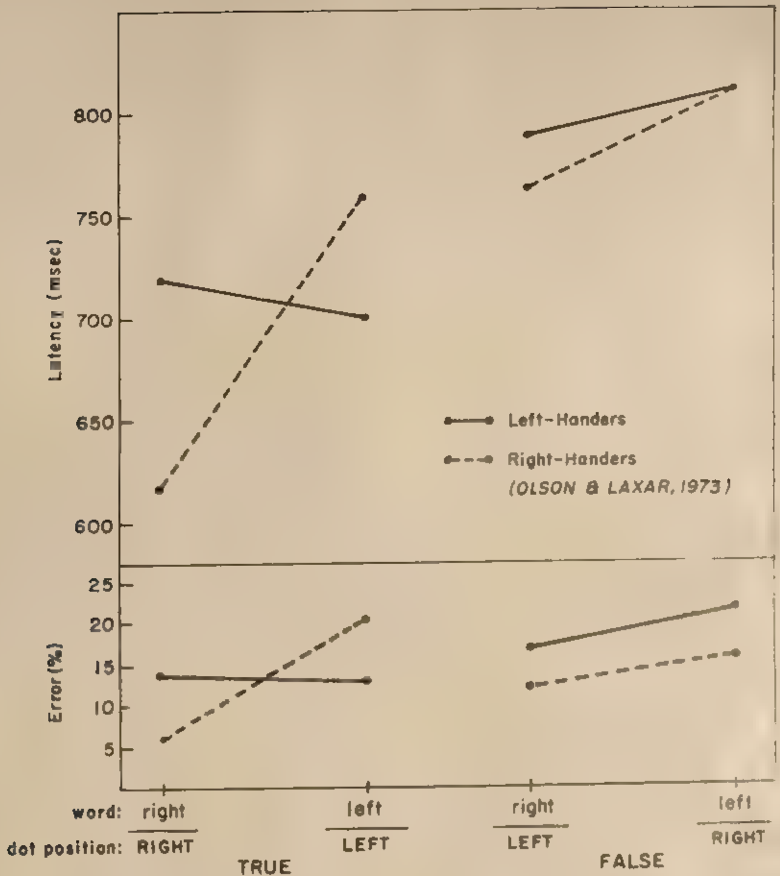


FIGURE 1. Mean reaction times for correct responses and error percentages for right- and left-handed *S*s.

The *S*s sat in a special experimental booth which contained the viewer for a three-field tachistoscope and a response panel at desk-top height immediately below the viewer. This panel contained two keys for *true* and *false* responses and a start button centered above the keys. The *S*s initiated each trial by pressing the start button with one of their middle fingers while their forefingers were positioned on the two response keys. The position of the *true* key was counterbalanced across *S*s. One second after they pushed the start button one of the four displays flashed on and *S*s responded *true* or *false* as quickly as possible. A fixation point was used to direct *S*s to look at the word first. On each trial *S*s were informed of their time and correctness.

The *S*s were 16 left-handed civilian and military personnel, 14 males and 2 females. Left-handedness was judged by asking each *S* which was her or his preferred hand. All *S*s wrote with their left hands.

Results and discussion. Mean latencies were computed for correct responses only for each Block × Display × *S* combination, yielding 16 means for each *S*. Since the performance of the 2 females did not differ from that of the males, sex of *S* was ignored in all of the analyses. The mean latencies and error rates for the four displays collapsed over blocks are shown in Figure 1. Comparable data from Experiment I of Olson and Laxar (1973) are also shown.

A repeated measures analysis of variance of the 256 means for the effects of stimulus word (RIGHT–LEFT), dot position (right–left) and test block (1–4) revealed the following significant effects: (*a*) *true* displays were verified more quickly than *false* ones, yielding a reliable interaction between RIGHT–LEFT and the positions right–left, F (1, 15) = 28.47, $p <$ .001; (*b*) *S*s became faster as the experiment progressed, F (3, 45) = 13.32, $p < .001$. The Dot Position × Blocks interaction approached significance, F (3, 45) = 2.34, $.05 < p < .10$. Analysis of errors revealed no significant F ratios, although the Stimulus Word × Dot Position interaction, F (1, 15) = 3.87, $.05 < p < .10$, and the main effect of blocks, F (3, 45) = 2.41, $.05 < p < .10$, approached significance.

As in the previous experiments with right-handers, stimulus-response (S–R) compatibility effects were very strong. These can be summarized as follows. For each *S* one of the *true* displays referred to the same side as the *true* response key, and this repre-

TABLE 1

SUMMARY OF SIGNIFICANT F RATIOS FOR ANALYSIS OF VARIANCE CONTRASTING LEFT- AND RIGHT-HANDERS

Source of variation	Primary effect	Interaction with handedness	MS for appropriate error term
Stimulus word	$F(1, 24) = 3.39$*	$F(1, 24) = 3.39$*	.0718
Position of dot	*ns*	$F(1, 24) = 4.39$**	.0272
Test blocks	$F(3, 72) = 45.83$***	$F(3, 72) = 3.09$**	.0131
Word × Dot (*true–false*)	$F(1, 24) = 49.09$***	*ns*	.0204

* $.05 < p < .10$.
** $p < .05$.
*** $p < .001$.

sented a compatible (C) relationship. The other *true* display represented an incompatible (IC) relationship. This was also true for the *false* displays. For each S the average difference between the correct reaction times for C and IC responses (IC − C) was computed for *true* and *false* displays separately. The mean difference of 171 msec. for *true* displays was significant, $t(15) = 5.52$, $p < .001$, while the difference of −16 msec. for *false* displays was not. By comparison, the right-handers of Olson and Laxar (1973) had a mean difference of 116 msec. for *true* displays and 17 msec. for *false* ones. An analysis of variance revealed that these S–R compatibility effects for left-handers did not interact with any of the effects revealed by the main analysis, while t tests showed that there was no difference between right- and left-handers.

A comparison of the results for right-handers and left-handers in Figure 1 reveals that their performance was quite different. These data were subjected to an unweighted-means analysis of variance with handedness as a between-Ss factor and stimulus word, dot position, and test block as within-Ss factors. There was no main effect for handedness. Table 1 summarizes the F ratios for the primary within-Ss sources of variation and their interactions with handedness. The only significant F ratio among the other higher order interactions was the triple Stimulus Word × Dot Position × Test Block interaction, $F(3, 72) = 2.93$, $p < .05$. These analyses confirm what can be seen in Figure 1. Of the three possibilities listed in the introduction, c appears to be supported. The absence of any higher order interactions either in the main analysis or in the secondary analysis for effects of S–R compatibility suggests that within this task there was no consistent pattern of trade-offs between right-normalized and left-normalized spatial models. However, no distinction between the absence of a reference direction and conflicting spatial models can be drawn from our evidence. Since our definition of left-handedness was whether S categorized himself as left- or right-handed, perhaps a more sensitive index of handedness or of hemispheric lateralization of function might offer greater insight into the left-hander's conceptualization of space.

The striking contrast between right-handers and left-handers on this task is consistent with the findings of other investigators. For instance, others have found that the performance of left-handers in tasks requiring spatial abilities is neither parallel to nor complementary to that of right-handers, but rather is intermediate (Luria, McKay, & Ferris, 1973). Disorders of spatial orientation are most acute in the right–left dimension (Corballis & Beale, 1970; Howard & Templeton, 1966), and there is evidence that right–left difficulties are much greater for left-handers (Gerhardt, 1959; Hécaen & de Ajuriaguerra, 1964). These facts along with the data reported here reinforce our earlier view that "the conceptualization of the right–left dimension is much more labile, flexible, and subject to disruption than that of the other two spatial dimensions [Olson & Laxar, 1973, p. 289]."

REFERENCES

CLARK, H. H. Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.
CORBALLIS, M. C., & BEALE, I. L. Bilateral symmetry and behavior. *Psychological Review*, 1970, **77**, 451–464.
GERHARDT, A. R. Left-handedness and laterality in pilots. *North Atlantic Treaty Organisation, Advisory Group for Aerospace Research and Development*, 1959, AGARDograph No. 30, 262–272.
HÉCAEN, H., & DE AJURIAGUERRA, J. *Left-handedness: Manual superiority and cerebral dominance.* New York: Grune & Stratton, 1964.
HOWARD, I. P., & TEMPLETON, W. B. *Human spatial orientation.* New York: Wiley, 1966.
LURIA, S. M., MCKAY, C. L., & FERRIS, S. H. Handedness and adaptation to visual distortions of size and distance. *Journal of Experimental Psychology*, 1973, **100**, 263–269.
OLSON, G. M., & LAXAR, K. Asymmetries in processing the terms "right" and "left." *Journal of Experimental Psychology*, 1973, **100**, 284–290.

(Received September 8, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1138–1140

SYLLABLE-DEPENDENT PRONUNCIATION LATENCIES IN NUMBER NAMING: A REPLICATION[1]

STUART T. KLAPP[2]

California State University, Hayward

Previous experiments have shown that the latency for the initiation of pronunciation of words and two-digit numbers depends on the number of syllables to be named. However, Henderson, Coltheart, and Woodhouse have recently shown that the presence of decades (e.g., 20, 30) was confounded with syllables in the number studies, and that this confounding could have accounted for some of the previous data. The present experiment is a successful replication of the naming latency effect for number stimuli controlling for this variable.

Several recent studies have reported that when single words or two-digit numbers are read aloud, the latency from stimulus onset until the *beginning* of overt vocalization increases as the number of syllables to be pronounced increases (Eriksen, Pollack, & Montague, 1970, Experiment III; Klapp, 1971, Experiment I; Klapp, Anderson, & Berrian, 1973, Experiment I). However, Henderson, Coltheart, and Woodhouse (1973) failed to obtain the syllable dependence in latency for naming *numbers*. The purpose of the present investigation was to reassess this phenomenon in light of the considerations raised by these investigators.

In their Experiment I, Henderson et al. (1973) used the same general procedure and material as were employed in the previous number studies (Eriksen et al., 1970; Klapp, 1971). The stimuli were two-digit numbers of two-, three-, and four-syllable pronunciation. The *S*s responded by pronouncing a number as quickly as possible after presentation, and time to initiate vocalization was determined as a function of the number of syllables to be pronounced. The results were ambiguous, with faster latencies for two syllables than for three or four syllables, but with no difference between the latencies for three and four syllables. These ambiguous results contrast with the monotonic increase in latency as a function of syllables reported under apparently identical conditions in the previous studies. However, in the Henderson et al. study there was no control condition to assess whether there were differences in the sensitivity of the equipment to the initial sounds associated with responses of different lengths. Klapp (1971) and Klapp et al. (1973) used a control in which *S*s examined the stimulus for a few seconds, and then pronounced it when a signal appeared. Latencies measured from this signal should be and were independent of the number of syllables. Since Henderson et al. did not provide these control data for their equipment, it is difficult to evaluate the extent to which the relative latencies assigned to the different classes of numbers may be an artifact of equipment activation. Further instability in the data could have resulted from the attempt to assess the relatively small effect of syllables from a somewhat limited number of observations—four trials per stimulus number for each of 15 *S*s, compared to eight trials for 12 *S*s in Klapp (1971

Although the results of the Henderson et al. (1973) Experiment I may be attributed to methodological factors, the results of their Experiment II are quite another matter. They pointed out that in the previous studies, including their own Experiment I, the two-syllable numbers were primarily decades (e.g. 30, 40, etc.), while decades did not appear among the three- and four-syllable numbers. They then showed that two-syllable decades were pronounced with shorter latencies than two-syllable teens (e.g., 13, 14, etc.). Since these stimuli have matched initial sounds, the lack of the above mentioned sound activation control is of no consequence in evaluating this particular comparison. Thus, confounding of decades could account for the previously reported shorter latencies for two-syllable numbers compared with three- and four-syllable numbers. Furthermore, most if not all of the overall effect of syllables reported previously by Eriksen et al. (1970) and by Henderson et al., Experiment I, was attributable to a shorter latency for two as compared to three and four syllables, with little or no effect of three compared to four syllables (although Klapp, 1971, did report a substantial three to four syllable effect). Therefore, Henderson et al. concluded that the previous reports of syllable-dependent latencies for naming numbers may well be attributed to the confounding of decades with syllables without the need to assume that response latency is related to number of syllables. Although these considerations are a problem only with number stimuli and do not apply to experiments demonstrating syllable-dependent response latencies for words (Klapp et al., 1973), it nevertheless appeared desirable to attempt a replication of this effect using new number stimuli not subject to this confounding.

Method. The stimuli were two-digit numbers ending in 7 (yielding four syllables in total pronuncia-

[1] This research was supported, in part, by the Small Grants Program of the Research Foundation of California State University, Hayward. The author wishes to thank Nancy Lankford, who gathered the data, and Alicia Klapp, who prepared the stimulus material.

[2] Request for reprints should be sent to Stuart Klapp, Department of Psychology, California State University, Hayward, California 94542.

tion, e.g., 27), in 6 (three syllables, e.g., 26) or in 4 (also three syllables, e.g., 24). According to the principle that vocalizations of more syllables require longer response latencies, the numbers ending in 7 should have longer latencies than the other numbers. However, which digit is used to terminate the number should not matter unless the two-syllable digit 7 is used. Therefore, numbers ending in 4 or 6 should yield latencies which are identical to each other while shorter than for numbers ending in 7. The particular digits 4 and 6 were selected to be distinctively different in vocalization, while equated for syllables. The first digits used in generating the two-digit stimulus numbers were 2, 3, 5, 8, and 9. These particular digits were selected to avoid teens and identical digits (e.g., 66). These five first digits were used with the three ending digits to generate a complete set of 15 stimulus numbers which were prepared on 35-mm. slides using Instantype numbers and projected to subtend a visual angle of 2° high by 3.7° wide. The numbers were presented in 15 blocks, each comprised of all 15 numbers in random order. A brief rest pause occurred after each block. The last 12 blocks were included in the analysis, while the early blocks were regarded as practice.

The *S*s were adult native speakers of English fulfilling a course requirement or serving for $2 pay. The *S*s were alternately assigned to control and experimental conditions until 8 *S*s had been assigned to each condition. Thereafter, all *S*s were assigned to the experimental condition, for a total of 13 *S*s in this condition. Data from an additional *S* in the experimental condition were rejected because 36% of the latencies exceeded 800 msec., including 20% of the total in excess of 1 sec. and several in excess of 2 sec.

The *S*s were seated in a sound-isolation chamber into which the numbers were projected on a rear-projection screen through double-glass sound insulation. The *E* was in an adjacent chamber containing apparatus which automatically provided the sequence of events for each trial. The response latency was measured by an electronic timer which was started when a photocell was activated by the onset of the image or signal light on the projection screen. This timer was stopped by a voice-activated trigger operated by a microphone placed in near contact with *S*'s mouth. Sensitivity was adjusted to the maximum possible, while avoiding frequent activation by breathing.

For the *control* condition, each trial lasted 7.9 sec., during which the stimulus number was present for 6 sec. After the number had been visible for 3 sec., a visual signal light appeared and the timer was started. The *S*s were instructed to pronounce the number as quickly as possible after the appearance of the signal, and the initial vocal sound caused the timer to stop. Any trials on which *S* pronounced the wrong number, corrected himself, or generated stray sounds were repeated at the end of the following block of 15 trials, and these data were included in the analysis.

TABLE 1

PRONUNCIATION LATENCY (IN MSEC.)

Condition	Stimuli: 3 syllable, Number ends in 6	3 syllable, Number ends in 4	3 syllable, $\bar{X}$	4 syllable, Number ends in 7
Control	267.9	269.9	268.9	269.0
Experimental	422.5	419.9	421.2	427.8

For the *experimental* condition, each trial lasted 6.9 sec., during which the number was presented simultaneously with the starting of the timer and remained visible for 3 sec. The *S*s were instructed to pronounce the number as quickly as possible after its appearance, and the initial vocal sound caused the timer to stop. Except for the change in sequence, the procedure was identical to that of the control condition.

Results. For the *control* condition, .4% of the trials had to be repeated due to number mispronunciation, while stray noise or equipment error necessitated repeating another 2.5% of the trials. On an additional .2% of the trials the latency exceeded 800 msec. These trials were assumed to represent inattention on the part of *S*, and the means were calculated without these data. Means were determined for each *S* for the sets of numbers ending in 7, 6, and 4, and the analysis was performed on these means. The results appear in Table 1. Latencies for the numbers with the three different endings did not differ significantly, $F\ (2,\ 14) < 1$, and the overall mean latency for three-syllable numbers was within .1 msec. of that for four-syllable numbers. Thus, the equipment was activated by the sounds of the different classes of numbers with apparently equal speed. This is certainly to be expected, since these sets of numbers were balanced with respect to initial sound.

For the *experimental* condition, the rates of repeated trials were 1.2% for number mispronunciation and 1.7% for stray noise or equipment error. For 1.6% of the trials the latency exceeded 800 msec. and, as in the control condition, the means were calculated excluding these data. There was no change in the direction of the data trend for effect of syllables as a result of discarding these long latencies. As is evident from Table 1, the mean latency was longer for four-syllable numbers than for either of the three-syllable conditions which, as expected, yielded approximately equal latencies. The three conditions differ significantly from each other, $F\ (2,\ 24) = 5.6$, $p < .01$, with both Newman-Keuls contrasts between four-syllable numbers ending in 7 compared to three-syllable numbers ending in 6 or in 4 significant beyond the $p < .05$ level. Four-syllable numbers yielded longer latencies than three-syllable numbers for 11 of the 13 *S*s ($p < .05$, sign test), with the remaining two *S*s showing

latencies which differed by less than 2 msec. between these conditions. Thus, the syllable effect for naming numbers has been replicated in an experiment which controls for the considerations raised by Henderson et al. (1973). The magnitude of the effect was 6.6 msec., which approximates the 5-msec. three- vs. four-syllable effect reported by Eriksen et al. (1970), but which is considerably less than the corresponding 14-msec. effect reported by Klapp (1971).

Discussion. The conclusion that syllable dependent response initiation latencies occur for numbers as well as for words suggests that this phenomenon is not due to perceptual analysis of the visual stimuli into syllabic units (Spoehr & Smith, 1973), since syllables are represented at the graphemic level only for words and not for the two-digit numbers. Although such perceptual analysis may occur for words, it cannot account for the syllable dependency in response latencies for numbers. Klapp et al. (1973) suggested an interpretation which assumes that a vocalization must be programmed before it can be emitted, with longer programming time for more syllables. This interpretation correctly predicts that the effect of syllables on pronunciation latency would occur for various modes of presentation, including numbers as well as words, and even extending to pictorial representations (Klapp et al., Experiment IV

REFERENCES

ERIKSEN, C., POLLACK, M., & MONTAGUE, W. [illegible]plicit speech: Mechanism in perceptual encoding? *Journal of Experimental Psychology*, 1970, **84**, 502–507.

HENDERSON, L., COLTHEART, M., & WOODHOUSE, D. [illegible] to find a syllabic-effect in number naming. *[illegible]nition*, 1973, **1**, 304–306.

KLAPP, S. Implicit speech inferred from response latencies in same-different decisions. *Journal of Experimental Psychology*, 1971, **91**, 262–267.

KLAPP, S., ANDERSON, W., & BERRIAN, R. Implicit speech in reading, reconsidered. *Journal of Experimental Psychology*, 1973, **100**, 368–374.

SPOEHR, K. T., & SMITH, [illegible] syllables in perceptual processing. *Cognitive Psyc[illegible]*, 197[illegible], **5**, 71–8[illegible]

(Received September 10

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1140–1142

SHORT-TERM MEMORY FOR SOUNDS AND WORDS[1]

EDWARD J. ROWE[2]

Memorial University of Newfoundland, St. John's, Newfoundland, Canada

RONALD P. PHILIPCHALK

St. Thomas University, Fredericton, New Brunswick, Canada

LESLIE J. CAKE[3]

Memorial University of Newfoundland, St. John's, Newfoundland, Canada

Memory for triads of words and meaningful environmental sounds was compared following a 20-sec. retention interval, during which *S*s shadowed passages of either poetry or music. Words were recalled better than sounds after music distraction, while the reverse was true following poetry distraction. This interaction suggests that short-term retention of sounds and words might be mediated by different memory processes.

Memory for common environmental sounds has only recently begun to be explored. Miller and Tanis (1971) compared recognition memory for long lists of printed words, spoken words, and sounds and found accuracy scores of 84%, 75%, and 69%, respectively, on a two-alternative forced-choice test. Thus, sounds were recognized at about the same level as spoken words, but at a much lower level than comparable performance reported by other investigators for recognition of meaningful pictures, where the accuracy level usually exceeds 90%. Lawrence and Banks (1973) improved upon the procedure employed by Miller and Tanis by using high-fidelity equipment and lengthening the presentation time for individual sounds. The overall recognition accuracy calculated from *S*s' confidence ratings was 85%, a figure noticeably higher than that reported by Miller and Tanis, but still somewhat below the level of picture recognition.

Philipchalk and Rowe (1971) asked whether memory for sounds vs. words would be differentially

[1] This research was supported by Grant A8580 to the first author from the National Research Council of Canada and by Grant 9401-55 from the Defence Research Board to the second author. The assistance of Lynn Wolff is gratefully acknowledged.

[2] Requests for reprints should be sent to Edward J. Rowe, Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada.

[3] Now at Queen's University.

affected by the requirement to recall the items in their presented order, as had been demonstrated for pictures vs. words in the visual modality. Their results showed comparable performance for sounds and words in free recall, but recall of sounds was significantly worse in a serial recall task. The fact that order information can be better retained for verbal material offers one potential distinction between the processes that mediate retention of sounds and words and raises the possibility of separate verbal and nonverbal auditory memory systems.

The experiment reported here represents a more direct test of the independent-systems hypothesis. Triads of sounds and words were presented for retention in the Brown–Peterson paradigm (Brown, 1958; Peterson & Peterson, 1959) with either verbal or nonverbal auditory distractor activity. The verbal distractor task (repeating passages of poetry aloud) was intended to disrupt verbal memory processes, and the nonverbal distractor (humming passages of music) to disrupt nonverbal memory. If sounds and words are encoded, stored, or retrieved via different systems, selective interference in recall of the two types of triads by the two types of distraction is expected, such that sounds should be recalled better than words following poetry distraction and words should be recalled better following music distraction.

Method. Twenty-four easily identifiable sounds, which ranged 1.4–4.4 sec. in length ($\bar{X} = 3.0$), were selected from a larger pool of items obtained from commercial sound effects recordings (see Philipchalk & Rowe, 1971). Two separate sets of eight triads were constructed and recorded on tape. The individual sounds within a triad were separated by 2 sec. or less of silence, giving a mean triad length of 10.4 and 10.5 sec., respectively, for the two sets. Parallel sets of triads of the dominant labels of the sounds were recorded in a male voice with the items being read at a rate of one every 3 sec. Thus the word triads were all 9 sec. long.

Four different versions of the experimental tape were drawn up for both sounds and words. One of two kinds of distractor material—music (M) or poetry (P)—was recorded on the tape immediately following each triad. The order of occurrence of the distractor material was PMMPMPPM for one version of each triad set and MPPMPMMP for the other. The M distraction consisted of the initial 20-sec. segments of instrumental melodies taken from various types of popular recordings. The P distraction included either complete poems or complete stanzas from poems, read in a female voice and timed to last 20 sec. The word "recall," spoken in a female voice, occurred at the end of each segment of distractor material, followed after 10 sec. by the words "stop writing," and after 3 more sec. by the next triad.

Most of the *S*s were paid volunteers from undergraduate psychology classes. Twenty *S*s each were assigned to the sounds and words conditions, with distractor type as a within-*S*s variable. The *S*s were tested individually or in pairs. If 2 *S*s were tested together, they were separated by a high cardboard shield that prevented them from seeing each other. They were instructed to listen carefully to each triad, to shadow the distractor material, and then to write down the names of the sounds or words from the triad on a prepared answer sheet when given the signal to recall. Ordered recall was not required. Practice in shadowing both music and poetry was provided prior to presentation of the experimental tape. Most *S*s experienced very little difficulty in repeating the poems aloud and humming along with the music segments, so no more than two trials on either the M or P practice material were necessary. The tapes were played through high-quality headphones, and *S*s' shadowing was recorded on a second tape recorder. Equal numbers of *S*s received the two different sets of triads in both the sounds and words conditions, but a slight error occurred in assigning *S*s to each distractor sequence for one set of triads in the words condition only. This resulted in 4 and 6 *S*s being run for each sequence instead of 5 each. Otherwise, the assignment of *S*s to list conditions was completely balanced.

After completion of the memory task, *S*s heard the same set of triads a second time with no distraction and were allowed 10 sec. for immediate recall of each set. The *S*s in the sounds condition were asked to use the same names for the sounds as they had used for recall in the first part of the experiment.

Results and discussion. To control for the possible effects of auditory recognition errors, the recall scores were conditionalized on performance in the second phase of the experiment. Thus, an item had to be reproduced correctly on both tests in order to be counted as correct. If *S* failed to recall an item on both tests, performance on the first test was scored out of the number of items recalled in that condition the second time and adjusted to a denominator of 12 for direct comparison with the other scores. This occurred for eight *S*s who heard the sounds and for one *S* who heard the words. No *S* missed more than one item on both tests.

The results of the second test were also used in scoring the protocols of the sounds group where *S*s' answers did not clearly correspond to the dominant label for a particular item. In these cases the sound was scored as correct if the same name was used on both tests. For both conditions, words which were minimally discrepant from a presented item (e.g., minor misspellings and substitution of plural for singular nouns) were counted as correct, as were homophones of presented items (e.g., *symbols* for *cymbals*). Incomplete recalls and intrusions from previous trials were scored as incorrect.

The recall scores (Table 1) were analyzed by a 2×2 analysis of variance, which yielded a significant main effect of distractor type, $F(1, 38) = 62.7$, and a significant Item $\times$ Distractor interaction, $F(1, 38) = 29.6$, both $ps < .001$. The interaction qualifies the distractor-type main effect, since only word recall was significantly affected by M vs. P distractor tasks. Of the 20 *S*s in the words condition, 19 had lower recall following P distraction, while the remaining *S* had equal scores for both M

TABLE 1

Mean Number of Items Correctly Recalled and Mean Time Spent in Shadowing as a Function of Music and Poetry Distractor Tasks

Type of distraction	Sounds		Words	
	Recall	Shadowing	Recall	Shadowing
Music	9.0	15.1	11.3	16.0
Poetry	8.1	11.4	6.9	11.0

and P triads. With the sounds, 10 *S*s had lower recall following P material, 7 showed the opposite effect, and 3 had equal scores. Comparison of sound vs. word recall within each distractor condition showed that words were recalled better under M distraction, t (38) = 3.71, $p < .001$, while sounds were recalled better under P distraction, t (38) = 1.77, $p < .05$ (one-tailed tests).

Because of the difficulties inherent in scoring *S*s' shadowing of music on an accuracy measure, shadowing performance for both types of distractor material was examined by measuring the total amount of time *S* was actively engaged in shadowing during the 20-sec. retention interval. The means are included in Table 1. Since the responses produced in speaking and humming are qualitatively different, meaningful comparisons between distractor types are not possible. It should be noted, however, that the lower scores produced in shadowing the poetry probably do not reflect a lower degree of accuracy than in the music conditions, as informal observation of the *S*s suggested that performance on the poetry was quite high. The difference between the words and sounds conditions for music shadowing was reliable, t (38) = 2.27, $p < .05$ (two-tailed test). This might be interpreted to mean that attempts to rehearse the words during the retention interval did not interfere with humming as much as attempts at rehearsing the sounds. The reverse is suggested by the poetry shadowing data, but the difference here was not significant, t (38) = 1.05.

The differential effect of music and poetry distractor tasks on recall of sounds and words constitutes evidence for separate verbal and nonverbal auditory short-term memory processes. The results do not allow a firm decision regarding the possible localization of the selective interference in terms of the tripartite division of memory into encoding, storage, and retrieval phases. However, the fact that the recall differences between sounds and words are reflected in the shadowing scores, especially for music shadowing, suggests that the distractor tasks acted to disrupt attempts to rehearse the triads. This conclusion is supported by the informal comments of *S*s, especially those in the words condition, who reported that it was much easier to "think about" the words while humming the melodies than while repeating the poems. Thus, it seems that the interference is at least partly attributable to processes involved in the maintenance of the items in storage through rehearsal, but possible additional effects on encoding or retrieval cannot of course be ruled out.

The form of the obtained interaction between type of triad and type of distraction offers a clue to the way in which sounds are represented in memory. Assume that sounds may be represented either directly in an auditory imagery store or indirectly via their verbal labels in a verbal short-term store, or both. If only the names of the sounds are held in memory, then music vs. poetry distractor tasks should have affected recall of sounds and words in the same way. This clearly did not happen. If the sounds were stored only as nonverbal images, we might expect that the nonverbal distractor task would be more disruptive than the verbal task. This also is not reflected in the data. The most plausible interpretation, then, is that the sounds were represented via both modes, so that retrieval of the image could be used to offset the effects of verbal interference and retrieval of the word label could help counteract the effects of nonverbal interference.

The independence of verbal and nonverbal auditory memory stores has been previously demonstrated by Deutsch (1970), who found that interpolation of additional tones between two comparison tones in a *same–different* judgment task impaired performance, whereas interpolated numbers did not. Furthermore, Crowder (1971) has shown that the interfering effect of a redundant suffix in immediate recall of auditory letter sequences is restricted to verbal suffixes, the effect disappearing when a nonverbal suffix (a buzzer) is used. Both of these findings have been interpreted as reflecting the operation of sensory or precategorical memory processes. The present data show that the distinction between verbal and nonverbal auditory systems is still viable when processing proceeds to the extent of "depth" characterized by more meaningful sounds and real word stimuli presented under less stringent temporal conditions.

REFERENCES

Brown, J. A. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

Crowder, R. G. Waiting for the stimulus suffix: Delay, decay rhythms, and readout in immediate memory. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 324–340.

Deutsch, D. Tones and numbers: Specificity of interference in immediate memory. *Science*, 1970, **168**, 1604–1605.

Lawrence, D. M., & Banks, W. P. Accuracy of recognition memory for common sounds. *Bulletin of the Psychonomic Society*, 1973, **1**, 298–300.

Miller, J. D., & Tanis, D. C. Recognition memory for common sounds. *Psychonomic Science*, 1971, **23**, 307–308.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

Philipchalk, R. P., & Rowe, E. J. Sequential and nonsequential memory for verbal and nonverbal auditory stimuli. *Journal of Experimental Psychology*, 1971, **91**, 341–343.

(Received September 11, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1143–1144

SHOCK-ASSOCIATED WORDS IN A NONATTENDED MESSAGE: A TEST FOR MOMENTARY AWARENESS

R. S. CORTEEN[1] AND D. DUNN

University of British Columbia

An experiment was carried out to test for momentary awareness of critical words in a nonattended message during shadowing. On only 1 out of 42 occasions when there was a significant galvanic skin response to critical nonattended words did an *S* signal awareness. This finding was taken as support for the hypothesis that there can be some semantic processing without awareness.

In a recent paper Corteen and Wood (1972) seemed to demonstrate that words that had been previously shock associated, and also words that had a class relationship with the shock-associated words, would give rise to a significant number of phasic skin conductance changes when they were embedded in a nonattended message. The concurrent attended message task involved shadowing a prose passage. The importance of the paper rests entirely on the claim that *S*s were not aware of the significant words in the nonattended channel. The idea of semantic processing without awareness suggests a wide range of interesting possibilities; but the study would be quite trivial if it could be demonstrated that *S*s were aware even momentarily of the significant words.

Corteen and Wood (1972, p. 311) employed a detailed and probing postexperimental interview procedure which certainly appears to indicate that, at the time of the interview, immediately following the experimental sequence, *S*s were not aware of having heard any of the critical words. This procedure, however, does not come to terms with a major criticism, which implies that *S*s might have been aware of the critical words at the time they were presented but that the interference produced by the shadowing task was so great as to produce complete forgetting, not only of the specific words but of the experience of having been aware of anything at all. This is an important and relevant criticism and the present experiment was designed in an effort to provide an answer to it.

The basic technique used in this experiment was a slightly modified form of the method used by Treisman and Riley (1969) where they instructed *S*s to stop shadowing at once and tap if they heard particular target letters in either shadowed or nonshadowed messages. The results obtained by Treisman and Riley indicate that the technique is workable and that detection in nonshadowed messages can be made quite successfully. In the present study the only concern was with responses to targets in the nonattended message, but beyond this and the substitution of a buzzer for ruler tapping, the method used was essentially the same.

Method. The method used was, up to a point, identical to that used by Corteen and Wood (1972). Three city names were presented aurally along with other words and the city names were paired with a mild but unpleasant electric shock. Phasic skin conductance changes were measured concurrently and were used to indicate the successful formation of shock association. The shock-associated city names, together with three city names that had not been included in the shock-association procedure and three nouns that had been in the original shock-association list, were now presented to one ear of the *S*, embedded in a list of randomly selected unrelated nouns. Prior to this, *S*s were given practice in shadowing a prose passage. They were now presented with a prose passage in the ear opposite to that being stimulated by the city names and other nouns and asked to shadow the passage. Skin conductance changes were recorded throughout.

All that has been described above is outlined in detail by Corteen and Wood (1972). The major change was that suggested by the Treisman and Riley (1969) paper. Prior to the dichotic listening sequence *S*s were shown a Morse key which activated a simple buzzer. They were asked to place the fingers of their free (right) hand on the key and they were told that if they heard any city names, in either ear, during the shadowing procedure, they were to stop shadowing immediately and press the buzzer. The importance of accurate shadowing was stressed but the need simultaneously to stop shadowing and press the buzzer upon hearing *any* city name was heavily emphasized.

A total of 19 experimental and 17 control *S*s were run. All *S*s were undergraduate students at the University of British Columbia. The control *S*s were exposed to exactly the same procedure as the experimental *S*s except that there was no prior shock association. The purpose of the control *S*s was to compare the incidence of buzzer pressing to city names when there had been no shock pairing. As it happened the control was unnecessary.

Results. The study confirmed the original findings of Corteen and Wood (1972) almost exactly. Indeed it would have been dismaying if it had not, as it was an almost exact replication. The findings concerning awareness appear to be quite unequivocal. The 19 experimental *S*s were each exposed

[1] Requests for reprints should be sent to R. S. Corteen, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada.

to six "critical" words, three shock-associated city names and three other city names. This gave, in all, 114 opportunities to press the buzzer. The buzzer was pressed only once. Naturally this produced a phasic skin conductance response but it was impossible to determine whether this response was to the word or to the process of buzzer pressing. There were 41 significant phasic conductance responses to the remaining 113 critical words and on none of these occasions was the buzzer pressed. Of these 41 responses, 24 (42.1%) were to shock-associated city names and 14 (29.8%) were to non-shock-associated city names. There was a 7.3% responding rate for shock-list control nouns and a 5.7% responding rate for non-shock-list control nouns. The autonomic responding rate to city names of the control *S*s was higher than the rate to control nouns (12.3% vs. 6.9%) but the difference was not significant. The size of a significant phasic response is defined by Corteen and Wood as a change in resistance of at least 1 kΩ occurring within 3 sec. There were two buzzer presses by control *S*s.

Discussion. The hypothesis being tested by this experiment was that the results obtained in the Corteen and Wood (1972) study could be explained if *S*s were momentarily aware of stimulus words in the nonattended channel, but that this awareness was of such a transitory nature that no vestige of it remained at the end of the experiment. The hypothesis was suggested, but only barely supported, by the findings of Norman (1969) and Glucksberg and Cowen (1970) that memory for nonattended items, when they are probed for, persists for only a few seconds. Neither of these two cited studies demonstrated or sought to demonstrate that the items are in conscious awareness prior to being probed for, but they do suggest the idea of momentary awareness with subsequent complete loss.

The findings reported here seem to invalidate the above hypothesis. They do not, of course, deny the findings of Norman (1969) or Glucksberg and Cowen (1970) because, of course, the items they report memory for may only come into awareness when there is specific interruption and probe. But the results do certainly support the position taken by Corteen and Wood (1972) that there is no awareness of certain words in a nonattended channel even though those words are capable of giving rise to a significant phasic conductance response.

It is certainly still possible to argue about the precise meaning of the word "aware"; but insofar as it might be defined as the ability of *S*s to indicate the presence of a stimulus at the time of its occurrence, then this study strongly suggests that there can be autonomic responding to words that are not in the awareness of *S*s at the time of the responding.

REFERENCES

Corteen, R. S., & Wood, B. Autonomic responses to shock-associated words in an unattended channel. *Journal of Experimental Psychology*, 1972, **94**, 308-313.

Glucksberg, S., & Cowen, G. N., Jr. Memory for nonattended auditory material. *Cognitive Psychology*, 1970, **1** 149-156.

Norman, D. A. Memory while shadowing *Quarterly Journal of Experimental Psychology*, 1969, **21**, 85-93.

Treisman, A. M., & Riley, J. G. A. Is selective attention selective perception or selective response: A further test. *Journal of Experimental Psychology*, 1969, 79, 27-[illegible]

Received September 15, 1973.

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1145–1147

RELATION BETWEEN SIMILARITY GROUPING AND PERIPHERAL DISCRIMINABILITY[1]

JACOB BECK[2]

University of Oregon

Four experiments studied the extrafoveal discriminability of a tilted T, an L, an X, and a sideways T from an upright T. In a patterned field in which an *S* did not know where to direct his attention, the discriminability of the figures corresponded to their effectiveness in producing similarity grouping in a pattern of upright Ts. In an empty visual field in which the brightness between a figure and a background was able to direct an *S* where to attend, the discriminability of the figures did not correspond to their effectiveness in producing similarity grouping. The results are interpreted as supporting the hypothesis that similarity grouping involves segregating a pattern into groups on the basis of extrafoveal difference signals that are responded to in parallel, or, if sequentially, very quickly and independently of focal attention.

Beck (1972) proposed that similarity grouping involves segregating a pattern into groups on the basis of extrafoveal stimulus differences that are readily discriminated when attention is distributed and not concentrated or focused. He reported that the similarity grouping produced by rotating some of the figures in a display corresponded to their extrafoveal discriminability in a patterned visual field in which an *S* did not know where to attend, but not in an empty visual field. This difference in the discriminability of figural differences in a patterned and in an empty visual field was interpreted as resulting from a narrowing or focusing of attention. In an empty visual field, the brightness difference between a figure and a background was able to direct an individual where to attend. The present experiments further compare the extrafoveal discriminability of a tilted T, an L, an X, and a sideways T from an upright T in a patterned and in an empty visual field. Beck (1966, 1967) reported that similarity grouping occurs strongly for a tilted T but not for an L, and X, or a sideways T in a pattern of upright Ts.

Experiment I. The patterned visual field consisted of four figures in the corners of an imaginary square 7.62 cm. on a side. Each target consisted of either four upright Ts in each of the four corners or three upright Ts in three corners and an L, a T rotated 45° clockwise (tilted T), or a T rotated 90° counterclockwise (sideways T) in the fourth corner. The lines composing the figures were always at right angles to each other. The figures were made out of black dry-transfer lines .04 cm. wide (approximately .04°), and .32 cm. long (approximately .3°) affixed to white paper. The targets were presented by means of a Gerbrands three-channel tachistoscope in which the preexposure and exposure fields were masked so that a target appeared in a circular field 18.89 cm. in diameter. The luminances of the white backgrounds of the preexposure and exposure fields were 17.13 cd/m². Following the presentation of a target, the preexposure field immediately reappeared. A black dot placed in the center of the preexposure field served as a fixation point. A dot was also placed in the center of the square array on each target and carefully aligned to coincide with the fixation point in the preexposure field. An *S* was instructed to fixate the dot in the center of the square array and told that this would enable him to perform optimally. The distances between the center dot on a target and the figures in the corners were 5.33 cm. (5° 7′ of arc).

The targets with an L, tilted T, and sideways T were presented in three separate sets. Each set consisted of four targets with an upright T in each of the four corners and four targets with an upright T in three corners, and a disparate figure in the fourth corner. An *S* was told before the presentation of each set which disparate figure would be presented in that set. He was instructed to report "same" for targets having all upright Ts and to identify the location of the disparate figure for targets having a disparate figure. An error was counted if an *S* reported a disparate figure when none was present, failed to report a disparate figure, or incorrectly reported the location of a disparate figure. Three of the 8 stimuli in a set were randomly added to a block of trials but were not analyzed (making 11 stimuli in all). The 11 stimuli in a set were presented twice to an *S* in a random order for a total of 22 presentations. The three sets were presented in a different order to each *S*. Thus, each *S* received 66 presentations in the experiment. Each *S* also received 66 presentations in a practice session before the experiment began. The practice period was used to determine an exposure duration that would give an overall percentage correct of between 70% and 75%. In Experiment I, the exposure durations ranged between 23 and 42 msec., with a mean of 30.3 msec.

[1] This research was supported by National Science Foundation Grant GB-24884. The paper was written while the author was a visiting member of the Computer Science Center at the University of Maryland. The preparation of the paper was supported in part by National Science Foundation Grant GJ-32258X.

[2] Request for reprints should be sent to Jacob Beck, who, during 1974, is with the Psychobiology Program, National Science Foundation, Washington, D. C. 20550.

TABLE 1

MEAN ERRORS AND STANDARD DEVIATIONS IN THE EXPERIMENTS

Experiment	*M*	*SD*
I: Patterned field ($N = 15$)		
Tilted T vs. T	3.1	1.6
L vs. T	4.9	1.5
Sideways T vs. T	6.4	1.6
II: Empty field ($N = 10$)		
Tilted T vs. T	4.3	1.9
L vs. T	3.5	2.2
Sideways T vs. T	5.0	1.5
III: Patterned field ($N = 15$)		
Tilted T vs. T	4.1	2.8
X vs. T	6.2	2.9
Sideways T vs. T	7.5	2.1
IV: Empty field ($N = 15$)		
Tilted T vs. T	2.9	1.6
X vs. T	3.0	1.7
Sideways T vs. T	5.6	1.4

Experiment II. In Experiment II, a single figure appeared on one corner of an imaginary 7.62-cm. square in an otherwise empty visual field. Three separate sets of stimuli were prepared. Each set consisted of four targets containing a single upright T in each of the four corners of the imaginary square and four targets containing either an L, a tilted T, or a sideways T in each of the four corners. An *S* was told before the presentation of each set which two figures (upright T and tilted T, for example) would be presented. An error was counted if an *S* incorrectly identified either a figure or the location of a figure. The procedure and experimental arrangement were the same as in Experiment I. In Experiment II, the exposure durations ranged between 11 and 35 msec., with a mean of 17.4 msec.

Experiments III and IV. In Experiment III, the targets were similar to those in Experiment I. Each target consisted of either four upright Ts or three upright Ts and an X, a tilted T, or a sideways T. In Experiment IV, the targets were similar to those in Experiment II. Each target consisted of a single upright T, tilted T, X, or sideways T on one corner of an imaginary square in an otherwise empty field. The stimulus sets, procedure, and experimental arrangement in Experiment III were the same as in Experiment I and those in Experiment IV the same as in Experiment II. The only differences were that the figures in Experiments III and IV were drawn in India ink and the vertical and horizontal lines of the exposure and preexposure fields were not masked. The exposure durations in Experiment III ranged between 10 and 20 msec., with a mean of 18 msec., and in Experiment IV ranged between 10 and 20 msec., with a mean of 15 msec.

Subjects. Four separate groups of *Ss* who had either normal vision or corrected-to-normal vision when wearing glasses served in the experiments. Fifteen *Ss* served in Experiments I, III, and IV and 10 *Ss* served in Experiment II.

Results. Table 1 presents the mean errors and standard deviations for the experiments. The error rate overall in Experiment I was [illegible]%. Statistical tests of the greater mean error for an L than for a tilted T, t (14) = 3.83, $p < .01$; for a sideways T than for a tilted T, t (14) = 6.49, $p <$ [illegible]; and for a sideways T than for an L, t (14) = 2.96, $p < .05$, were significant. The error rate overall in Experiment III was 37%. Statistical tests of the greater mean error for an X than for a tilted T, t (14) = 2.54, $p < .05$, and for a sideways T than for tilted T, t (14) = 4.21, $p < .01$, were significant. The greater mean error for a sideways T than for an X failed to be significant, t (14) = 1.60, $p > .1$. Thus, the results show that, in a patterned visual field, *Ss* made statistically significantly more errors in discriminating an X, L, and sideways T from an upright T than in discriminating a tilted T from an upright T.

The error rate overall in Experiment II was 27%. Statistical tests of the greater mean error for a sideways T than for a tilted T, t (9) = 1.07, $p > .2$; for a tilted T than for an L, t (9) = 2.22, $p > .05$; and for a sideways T than for an L, t (9) = 1.90, $p > .05$, were not significant.[3] The error rate overall in Experiment IV was 24%. A statistical test of the greater mean error for an X than for a tilted T, t (14) = .26, $p > .5$, was not significant. The greater mean error for a sideways T than for a tilted T, t (14) = 6.04, $p < .01$, and for a sideways T than for an X, t (14) = 5.71, $p < .01$, were significant. When a brightness difference between a figure and a background directed an *S* where to attend, the discriminability of an X and of an L from an upright T was similar to that of a tilted T. In an empty visual field, the discriminability of the figures, as hypothesized, failed to correspond to their effectiveness in producing similarity grouping.[4]

Discussion. The results are consistent with the hypothesis that the effectiveness of stimulus differences in producing similarity grouping correspond to their peripheral discriminability in a patterned visual field in which an *S* does not know where to attend. Beck (1966, 1967) found that Ls and sideways Ts in a pattern of upright Ts grouped more poorly than did tilted Ts. He also found that,

[3] The acceptance of the null hypoth[illegible]poses the problem of a Type II error. In Experi[illegible] an N of 10 increased this possibility. It is, therefo[illegible]nt to point out that the relative discriminability of a[illegible]nd of a tilted T in Experiment II was the opposite of t[illegible]n Experiment I. Though just missing significance, the mean error for an L was less than that for a tilted T in Experiment II. In contrast, in Experiment I, the mean error for an L was significantly greater than for a tilted T. Beck and Ambler (1972, 1973) also found that the peripheral discriminability of an L from an upright T was greater than or equal to that of a tilted T when attention was focused, but that the discriminability of an L became worse than that of a tilted T when attention was distributed over a patterned field. What is suggested is that focused attention increases the sensitivity of the visual system to the variable of line arrangement in the periphery. Information about line arrangement would facilitate discriminating an L from an upright T but would, if anything, because of their similarity in line arrangement interfere with discriminating a tilted T from an upright T.

[4] In Experiment IV, a tilted T was more discriminable from an upright T than a sideways T, but in Experiment II, it was not. A possible explanation is that the vertical and horizontal lines of the exposure field visible in Experiment IV served as cues for the tilt of a figure's lines. It is possible that masking the vertical and horizontal lines in Experiment II reduced the discriminability between a tilted T and an upright T.

though an X had the same slopes of lines as a tilted T, Xs grouped less readily than tilted Ts and similar to sideways Ts in a pattern of upright Ts.

An interpretation of the relation between similarity grouping and peripheral discriminability under uncertainty in a patterned field can be given by assuming that both are based upon stimulus differences computed from the figural properties that are responded to "quickly" and independently of focal attention (Beck & Ambler, 1972). What is proposed is that when attention is distributed over a patterned field, a tilted T is discriminated from an upright T on the basis of a parallel or, if sequential, almost simultaneous readout of differences in their line slopes from a visual information store (Beck & Ambler, 1973). The presence of confusable vertical and horizontal lines is assumed to interfere with a "quick" readout of differences in line arrangement. Thus, an L and a sideways T group poorly and are not discriminable with distributed attention in a pattern of upright Ts. Selective attention, however, can increase an *S*'s sensitivity to peripherally presented differences in line arrangement (Beck & Ambler, 1972, 1973). When a single figure is presented, the brightness difference between a figure and background directs an *S* where to attend and thereby improves the discriminability of an L, sideways T, and X.

The present results suggest that the Xs failed to group as well as the tilted Ts in a pattern of upright Ts because the central symmetry of an X causes it to readily lose discriminability outside of the fovea when it subtends a small visual angle. Under these conditions, the grouping of Xs in a pattern of upright Ts is weak. However, when the lines of an X are made larger, Xs grouped as well as tilted Ts (Beck, 1973).

REFERENCES

Beck, J. Perceptual grouping produced by changes in orientation and shape. *Science*, 1966, **154**, 538–540.

Beck, J. Perceptual grouping produced by line figures. *Perception & Psychophysics*, 1967, **2**, 491–495.

Beck, J. Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology*, 1972, **85**, 1–19.

Beck, J. *Similarity grouping and peripheral discriminability under uncertainty II.* (Tech. Rep. 233) College Park: University of Maryland Computer Science Center, 1973.

Beck, J., & Ambler, B. Discriminability of differences in line slope and in line arrangement as a function of mask delay. *Perception & Psychophysics*, 1972, **12**, 33–38.

Beck, J., & Ambler, B. The effects of concentrated and distributed attention on peripheral acuity. *Perception & Psychophysics*, 1973, **14**, 225–230.

(Received September 17, 1973)

Journal of Experimental Psychology
1974, Vol. 102, No. 6, 1147–1149

CATEGORY SIMILARITY AND RETROACTIVE INHIBITION IN FREE RECALL [1]

ISABEL M. BIRNBAUM [2]
University of California, Irvine

Two successive lists of categorized words were presented for free-recall learning. Half of the categories appeared on both lists (repeated categories) and half appeared on only one of the lists (nonrepeated categories). A standard control group learned one list. On the final test of free recall of both lists, different patterns of first-list retention loss were seen for repeated and nonrepeated categories. There was a loss of words within categories in the former case, and a loss of whole categories in the latter.

Recent studies indicate that retroactive inhibition (RI) in free-recall learning is an increasing function of similarity between the category composition of List 1 and List 2. Shuell (1968) found that RI was greater when different words from the same category appeared on both lists than when completely different categories were used on the first and second lists. With a mixed-list design where half of the List 1 categories were repeated (different words from the List 1 category appeared on List 2) and half were nonrepeated (no words from the List 1 category appeared on List 2), there was greater retention loss for words in the repeated categories (Winograd, 1968).

The present experiment was designed to investigate recovery of List 1 repeated and nonrepeated categories over a 10-min. retention interval following

[1] The author wishes to thank Karen Krueger and Joellen Hartley for their valuable assistance at every stage of the research.

[2] Requests for reprints should be sent to Isabel M. Birnbaum, School of Social Sciences, University of California, Irvine, California 92664.

TABLE 1

MEAN NUMBER OF WORDS RECALLED AND MEAN NUMBER OF WORDS RECALLED PER CATEGORY FOR REPEATED (R) AND NONREPEATED (N) CATEGORIES

Data	Control[a]		E_1		E_2	
	R	N	R	N	R	N
Words	14.05	15.22	11.11	11.50	12.44	11.66
Words per category	3.89	3.85	2.94	3.39	3.23	3.49

Note. E_1 = experimental group with a predominance of study trials and E_2 = experimental group with a predominance of test trials in interpolated learning.
[a] R vs. N was a dummy variable in the control group.

two types of interpolated learning (IL); however, chance differences in original learning (OL) in the immediate recall groups precluded the planned comparisons. The remaining groups provided a modified replication and extension of the original phenomenon, and results in these groups are the basis for the present report.

Method. The *S*s were 54 students at the University of California, Irvine, whose participation in the experiment contributed to the fulfillment of a course requirement. The *S*s were run individually and were assigned to groups in randomized blocks in order of appearance at the laboratory.

Lists for OL and IL contained 48 words, 6 words from each of 8 categories. A repeated category appeared on the OL and IL lists and the specific words from the category were different on each list. A nonrepeated category appeared only on the OL list or the IL list. Three different 8-category lists were constructed from a pool of 12 categories with 12 words in each category, using the design described by Winograd (1968). Words were selected from the category norms of Battig and Montague (1969) and two subsets of 6 words from each category were balanced for mean frequency of occurrence in response to the category name. Each category was a repeated category one-third of the time, and any pair of lists contained 4 repeated and 4 nonrepeated categories. If the three lists are designated A, B, and C, then one-sixth of the *S*s in each group learned each of these OL–IL sequences: AB, AC, BA, BC, CA, and CB. Thus, each category and each subset of 6 items within the category was used equally often as a repeated and a nonrepeated category in OL and in IL.

Each list was tape recorded in four different random orders with the restriction that no words from the same category appear in adjacent positions. The words were recorded at a 1-sec. rate and three of the orders were used equally often as starting orders in OL and IL. All *S*s received four alternating study and test trials on List 1. The duration of each trial was 48 sec. and there was a 12-sec. intertrial interval. Three groups were differentiated on the basis of the interpolated task. The control (C) group worked on a paper-and-pencil numerical reasoning task throughout the retention interval. Two experimental (E_1 and E_2) groups received 12 trials of IL and then worked on the numerical reasoning task for 10 min. Group E_1 received a predominance of study trials during IL: each four-trial cycle consisted of three successive study trials followed by a single test trial. Group E_2 received a predominance of test trials during IL: each four-trial cycle consisted of a single study trial followed by three successive test trials. In all groups, the final recall test began 26.5 min. after the last test trial on List 1. The *S*s in Groups E_1 and E_2 were asked to recall the words from Lists 1 and 2 in any order; recall was oral and was terminated after 4 min. had elapsed. The *S*s in Group C were given 2 min. for oral recall of the words from List 1 in any order.

Results and discussion. The mean number of words recalled on the last test trial of OL was 24.8, 22.6, and 24.5 in Groups C, E_1, and E_2, respectively, F (2, 51) < 1.00. There was no significant difference in recall of words from to-be-repeated and to-be-nonrepeated categories, F (2, 51) $= 1.22$, $p > .25$.

Shuell (1968) used a between-*S*s design and found consistently poorer performance in IL when words from the same category appeared on Lists 1 and 2 than when new categories appeared on List 2. In the present experiment, performance on repeated and nonrepeated categories in the IL mixed list showed the same pattern. Three-way analyses of variance of recall in IL (Groups $\times$ Cycles $\times$ Type of Category) either on the first test trial of each cycle or the fourth trial of each cycle (a test trial for both groups) showed that fewer words (ps $< .05$) and a smaller number of words per category (ps $<$ .01) were recalled from repeated categories than from nonrepeated categories during IL. There was improvement over cycles in IL and Group E_1 showed consistently higher recall than Group E_2 (ps $< .01$).

First-list recall of words from repeated and nonrepeated categories in each group is shown in Table 1. The designation of words as repeated and nonrepeated in Group C was a dummy variable. Analyses of variance of absolute recall scores on the final test or loss scores (recall on Trial 4 of OL minus recall on the final test) yielded essentially the same results. There was a significant difference between groups in loss scores based on number of words recalled and number of words recalled per category, F (2, 51) $= 7.12$, $p < .01$, and F (2, 51) $= 5.09$, $p < .05$, respectively, confirming the phenomenon of RI in free-recall learning of categorized lists. There was no significant difference between E_1 and E_2 on either dependent variable (ps $> .10$) suggesting that first-list retention loss is not modified by the pattern of study and test trials in IL when the number of trials is held constant. There was greater RI for words recalled per category in repeated than nonrepeated categories, F (2, 51) $=$ 7.96, $p < .01$, for the Groups $\times$ Type of Category interaction, and there was no significant RI for nonrepeated categories alone, F (2, 51) $= 1.17$, $p > .25$. The results are consistent with those of

Tulving and Psotka (1971) who found no significant reduction in first-list within-category recall when successive free-recall lists contained different categories.

Winograd (1968) found that on the final test of recall, List 1 nonrepeated categories were less likely to be retrieved than repeated categories but more items within a recalled nonrepeated category were accessible. The interplay of these factors yielded an approximately equivalent total number of items recalled from repeated and nonrepeated categories, and the present results in the experimental groups confirm Winograd's findings. Category recall, defined as recall of at least one item from the category, was significantly higher for repeated than nonrepeated categories, F (1, 34) = 13.28, p < .01. It appears that the probability of recall of at least one word from a category was increased by the greater frequency and/or recency with which the category had been encountered. However, when different words from the same category were included in OL and IL there was a reduction in first-list within-category recall; given that a category was recalled, a significantly greater number of words from the category were recalled in nonrepeated than repeated categories, F (1, 34) = 8.68, p < .01. As in Winograd's study, there was no significant difference in the total number of words recalled from nonrepeated and repeated categories (F < 1.00). In contrast, Shuell (1968) found lower absolute recall as well as fewer words recalled per category when the same categories appeared on the first and second lists. The increase in category recall did not compensate for the reduction in within-category recall. On the basis of the present data and the data reported by Shuell and Winograd, it is clear that the occurrence of words from the same conceptual categories in two successive lists enhances recall of the categories on the final free-recall test and reduces the recall of first-list words within each category.

Given the large number of procedural differences between Shuell's (1968) and the present study, the similarity of results is striking. In addition to a smaller number and different pattern of trials in IL, Shuell used an unmixed-list design: Either all or none of the categories on List 1 were carried over to List 2. Thus, the overall contextual similarity between OL and IL (in terms of category composition) was higher when categories were repeated (same group) than when categories were not repeated (different group). In the present study, the overall contextual similarity of the interpolated list was somewhat similar to the original list for repeated or nonrepeated categories. Yet in both studies, significant within-category retention loss was found only for repeated categories. Nonrepeated categories appear to be highly resistant to interference, showing no reduction in within-category recall whether the categorical composition of the second list is completely changed (as in Shuell's study) or only partially changed (as in the present study). It will be of interest to determine whether a nonrepeated category remains impervious to interference when all but one category is carried over from OL to IL. In view of previous results on RI and contextual similarity, it seems reasonable to assume that within-category retention loss in nonrepeated categories will be a function of the similarity of the remainder of the interpolated list to the original list.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt. 2).

Shuell, T. J. Retroactive inhibition in free-recall learning of categorized lists. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 797-805.

Tulving, E., & Psotka, J. Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology*, 1971, **87**, 1-8.

Winograd, E. List differentiation, recall, and category similarity. *Journal of Experimental Psychology*, 1968, **78**, 510-515.

(Received October 9, 1973)

Special Announcement

Each year many of the manuscripts submitted to the *Journal* are reviewed by referees whose names do not appear on the Masthead of the *Journal* as regular Consulting Editors. Of the 805 manuscripts received in 1973, about 211 were reviewed by 73 such consultants. The names of 15 of these psychologists appear on the 1974 Masthead. The Editors of the *Journal* would like to express their appreciation for the services of all of these reviewers. They ably served the authors, editors, and readers of the *Journal* by their thorough and competent reviews. Their reviews were of the same high quality that we expect and get from our regular Consulting Editors, so that they helped to improve the authors' scientific communication and, in certain instances, aided the authors in improving their research methods. And so, the Editors wish to thank the reviewers listed below.

Fred Attneave
Richard J. Barclay
David J. Barker
Ian Begg
Ina McD. Bilodeau*
Isabel M. Birnbaum*
Elizabeth L. Bjork
John D. Bransford
Myron L. Braunstein
Fergus I. M. Craik*
Dwight W. Curtis
David H. Dodd
Robert Fox*
Wendell R. Garner
Sam Glucksberg
James G. Greeno
Anthony G. Greenwald*
G. Robert Grice
Sherman L. Guth
John C. Hay
Joseph B. Hellige*
James V. Hinrichs
Betty J. House*
Lloyd G. Humphreys
Peder J. Johnson*
Lawrence E. Jones
Lyle B. Jones
Mari R. Jones
James F. Juola
William E. Kappauf*
Steven B. Keele
Michael D. Klitzner
Marvin Levine
Charles Lewis
Alvin M. Liberman
Gregory R. Lockhead*
Elizabeth F. Loftus*
Lola L. Lopes
Steven F. Maier*
Dominic W. Massaro
Dorothy E. McAllister
Wallace R. McAllister*
Richard M. McFall
William E. Montague
Paul Neuman
Robert G. Pachella
Allen Parducci
James W. Pellegrino
Peter G. Polson
David C. Raskin*
Frank Restle
Russell Revlis*
Lance Rips
Peter H. Schiller*
Robert W. Sekuler*
James C. Shanteau
Peter D. Shaw
Wayne Shebilski
Richard M. Shiffrin
Harvey G. Shulman*
Arnold M. Small, Jr.
Edward E. Smith*
Kirk H. Smith*
Terry J. Spencer*
George E. Stelmach
John E. Taplin*
Willard R. Thurlow*
Michael T. Turvey*
Thomas S. Wallsten*
Thomas D. Wickens
Lucinda Wilder
Robert C. Williges
Eugene Winograd*

* Reviewed more than one manuscript.

INDEX

CONTENTS—(*Continued*)